NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

VOICE OVER INTERNET PROTOCOL TESTBED DESIGN FOR NON-INTRUSIVE, OBJECTIVE VOICE QUALITY ASSESSMENT

by

David L. Manka

September 2007

Thesis Advisor:	Murali Tummala
Co-Advisor:	John C. McEachen

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2007	Master's Thesis

4. TITLE AND SUBTITLE Voice Over Internet Protocol Testbed Design for Non-Intrusive, Objective Voice Quality Assessment

5. FUNDING NUMBERS

6. AUTHOR(S) David L. Manka

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

Voice over Internet Protocol (VoIP) is an emerging technology with the potential to assist the United States Marine Corps in solving communication challenges stemming from modern operational concepts. This thesis conducts a review of VoIP standards and develops an H.323-based testbed for the study of tactical wireless VoIP performance. Methods of collecting and presenting voice quality parameters in packet-based networks are explored. Incorporation of an Adtech SX/14 Data Channel Simulator provides user control of a SONET-simulated wireless channel. Experiments quantify the effect of channel injected error rate on received voice traffic. Plots are generated to illustrate the relationship between channel error rate, packet loss, and the listening quality mean opinion score. Experimental results are extended by incorporating E-model delay considerations. Commercial voice recognition software is successfully used to measure the impact of the channel on speech intelligibility. The experiments and analysis conducted provide a cost effective approach to non-intrusive, objective voice quality assessment.

14. SUBJECT TERMS
Voice over Internet Protocol, VoIP, Voice quality, MOS, Wireless communications

15. NUMBER OF PAGES
117

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

VOICE OVER INTERNET PROTOCOL TESTBED DESIGN FOR NON-INTRUSIVE, OBJECTIVE VOICE QUALITY ASSESSMENT

David L. Manka
Captain, United States Marine Corps
B.S., United States Naval Academy, 1998

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 2007

Author: David L. Manka

Approved by: Professor Murali Tummala
Thesis Advisor

Professor John C. McEachen
Co-Advisor

Professor Jeffery B. Knorr
Chairman, Department of Electrical and Computer Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Voice over Internet Protocol (VoIP) is an emerging technology with the potential to assist the United States Marine Corps in solving communication challenges stemming from modern operational concepts. This thesis conducts a review of VoIP standards and develops an H.323-based testbed for the study of tactical wireless VoIP performance. Methods of collecting and presenting voice quality parameters in packet-based networks are explored. Incorporation of an Adtech SX/14 Data Channel Simulator provides user control of a SONET-simulated wireless channel. Experiments quantify the effect of channel injected error rate on received voice traffic. Plots are generated to illustrate the relationship between channel error rate, packet loss, and the listening quality mean opinion score. Experimental results are extended by incorporating E-model delay considerations. Commercial voice recognition software is successfully used to measure the impact of the channel on speech intelligibility. The experiments and analysis conducted provide a cost effective approach to non-intrusive, objective voice quality assessment.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The evolution of digital technologies in the voice communications market presents new opportunities for organizations to achieve economic and performance savings. Circuit switched networks are being replaced by more efficient packet-based designs. As these improved networks permeate voice communications, organizations combining voice and data onto a common platform can reduce management and equipment costs.

Voice over Internet Protocol (VoIP) is one of the applications driving the trend towards converged packet-based networks. VoIP has enjoyed success in enterprise-level deployments of civilian and military facilities throughout the globe. Extending the reach of VoIP applications to the tactical military environment will assist in the reduction of a unit's logistics footprint. Administering a single converged network also allows the military to train a reduced variety of occupational specialties for maintenance needs. Among tactical units, wireless enabled VoIP would also facilitate operations in areas of reduced or damaged telecommunications infrastructure. The United States Marine Corps' vision for greater dispersion across the battlespace supports the demand for innovative communications solutions. Mobile wireless capabilities required for tactical actions offer less predictable performance when compared to a fixed, wired network design. Theses factors provide the motivation for this thesis research.

The objectives of this thesis are divided among two principal tasks. First, this research develops a flexible, scalable VoIP testbed based on the H.323 standard. Using the Adtech SX/14 Data Channel Simulator, the experimental VoIP network provides user control of a SONET-based representation of the wireless channel. The effect of channel bit error injection is monitored for effects on packet loss, received voice file listening quality mean opinion score (MOS-LQK), and remaining speech metrics. Second, this thesis investigates methods of collecting and presenting voice quality parameters in packet-based networks. Emphasis is placed on non-intrusive, objective voice quality assessment methods that accommodate dynamic testbed topologies. Additionally, predicted delay effects are quantified, using the E-model, and presented as an extension to experimental results.

VoIP implementation is primarily divided among two competing standards for call signaling and control. Session Initiation Protocol (SIP), a product of the Internet Engineering Task Force (IETF), uses a series of text-based message exchanges to control audio, video, and data transfer sessions. SIP's control features are similar to the approach developed within Hypertext Transfer Protocol (HTTP). In contrast, H.323 has emerged from sources related to more traditional telephone standards, the International Telecommunications Union (ITU). While IETF and ITU feature disparate VoIP call control and signaling structures, both standards use Real Time Protocol (RTP) encapsulated within a User Datagram Protocol (UDP) packet for the end-to-end transport of sampled voice data. The unreliable nature of this form of telephony imposes network effects on the performance of voice related services.

Degradation of voice quality in any communications system can be broken into a set of additive impairment factors: echo, delay, and clarity. Once the impact of network effects is quantified among these subdivided metrics, the cumulative impact on voice quality is reported according to ITU-defined standards for subjective, objective, or predictive testing methods. Subjective testing requires a costly and time consuming direct interaction between human subjects for experimentation. In an effort to maximize scalability and flexibility of the testbed, this thesis explores ITU methods of objective and predictive voice quality assessment. Results from testbed techniques are presented in a MOS-LQK format, where 1 is bad and 5 is excellent in voice quality. Results from objective and predictive methods are highly correlated to scores obtained through subjective tests. Measurements can be obtained from a single receiver terminal without direct input from uncorrupted reference file transmission. This non-intrusive, single-ended structure provides added testbed flexibility for future research efforts.

The testbed design developed in this thesis incorporates Cisco 2851 and 7200 routers to replicate a two-site, distributed call processing model. Each site conducts independent call processing using Cisco 7825 Media Convergence Servers (MCS) running CallManager 5.0. A web-based configuration utility allows testbed users to set the network codec and manage devices registered to the CallManager software. The Adtech SX/14, positioned between each Cisco 7200 router, provides wireless channel simulation between CallManager clusters. Reference files for voice experimentation are maintained on a MCS for selective playback initiated through a call hold sequence. Network packet traffic analysis, VoIP call recording, and speech recognition are provided by Wireshark 0.99.5, Cain and Abel v4.9.1, and Dragon NaturallySpeaking software tools, respectively.

Experimentation shows valid Gaussian distributed random error rates can range from 1×10^{-12} to 2×10^{-5} error/bit. Errors injected at a rate greater than 2×10^{-5} produce link failure between the Cisco 7200 routers. Each codec suffered a corresponding decline in MOS-LQK as channel errors increased. Experiments achieved an approximate MOS-LQK range of 4.5 to 3.5 for G.711 and 3.7 to 3.5 for G.729. Except for the most severe error rate available to the testbed, G.711 provided superior MOS-LQK performance for all data points. Analysis reveals a decrease in MOS-LQK consistent with the increase in lost packets for both codecs. G.729 tests suffered less overall packet loss compared to G.711 runs. Remaining speech computation revealed an important distinction between the perception of VoIP listing quality, measured by MOS-LQK and intelligibility. Files captured at lower MOS-LQK scores still managed to deliver near perfect remaining speech results. G.729 with a MOS-LQK of 3.7 provided superior comprehension to the listener when compared to G.711. Experimental results were extended by analytically incorporating E-model predicted delay effects, which estimate decreased user VoIP quality satisfaction related to satellite links. Military applications may favor the benefit of voice connectivity in remote regions over the impairment effect of geosynchronous satellite delay.

The objectives of this thesis were explored and successfully addressed. Military deployment of wireless VoIP solutions in a tactical environment requires a dedicated platform for experimentation. A reconfigurable H.323-based VoIP testbed was developed and studied using ITU recommended voice quality measurement techniques. Objective, non-intrusive voice quality measurement methods were introduced for future research efforts.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to extend the most sincere appreciation to my advisors, Professors Murali Tummala and John McEachen, for their dedicated instruction and editing throughout the thesis process. I would also like to thank Mr. Bob Broadston and Lieutenant Nikolaos Tiantioukas for their assistance in making my laboratory setup a reality.

Finally, I would like to acknowledge the unwavering support of family and friends, especially my loving wife, Autumn. She has been a constant source of strength through deployments, cross-country moves, and long hours of study.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The past two decades have witnessed a transformation in the technologies used to provide commercial voice services. Traditional telecommunications, previously divided among broadcast and point-to-point applications, are rapidly converging to a unified model of diverse applications that promise to revolutionize the fractured concepts of multimedia exchange. Just as cable companies challenged the notion of television, the Internet based transfer of voice traffic is poised to revolutionize modern telephony.

The evolution of cellular phone technology offers a case study on the impact of disruptive inventions of the last century. Over the course of four decades, cellular phone subscribers have emerged as the dominant population in the world telephone market [1]. The next generation of cellular technology plans to upgrade mobile subscribers to an all packet-based network. This surge in development has largely been fueled by the associated transformation of wireline services incorporating another disruptive technology, Voice over Internet Protocol (VoIP).

When VoIP pioneers started plugging microphones into their computers in the 1990s, the economic impact shocked the telecommunications industry. Near ubiquitous broadband Internet access in major markets allowed reasonable quality voice connections directly between PC terminals. PC-to-PC calling suddenly offered a cheap innovative alternative to regular phone service. These early toll bypass exchanges lacked well accepted implementation standards and reliability. In contrast, the international standards of today make VoIP a dependable telephony option across the globe. Interconnections with the Public Switched Telephone Network (PSTN) have extended the scope and flexibility of VoIP. Faced with the prospect of losing millions of subscribers, telephony providers now compete for consumers with bundled data, video, and voice packages that often utilize VoIP technology [2].

The transformation of civilian communications continues to shape and influence military voice services. VoIP joins the growing collection of satellite and terrestrial based tools the military relies on for command, control, communications, computers and intelligence (C4I). These links are critical to the vision outlined in [3]. Publication of [3] officially updated and unified the core operational capabilities described by Operational Maneuver from the Sea (OMFTS), expeditionary maneuver warfare (EMW), and Distributed Operations (DO). These operating philosophies, collectively referred to as the Coherent Concepts, place strenuous demands on C4I capabilities. VoIP is part of a broad solution to growing military demands for multimedia capability in expeditionary environments.

Cost, capacity, and performance limitations continually challenge our efforts to network expanding battlespace geometry. Applications joining the existing architecture face increased competition for bandwidth allocations. At the tactical level, factors are exacerbated by link distance, mobility, and hostile environments. Efforts to improve network capacity must be complimented by a focus on the efficient use of existing resources. Advanced wireless technologies combined with VoIP provide comprehensive solutions to many networking hurdles. Figure 1 provides an illustration of potential network links augmented by IEEE 802.11 and 802.16 capabilities.

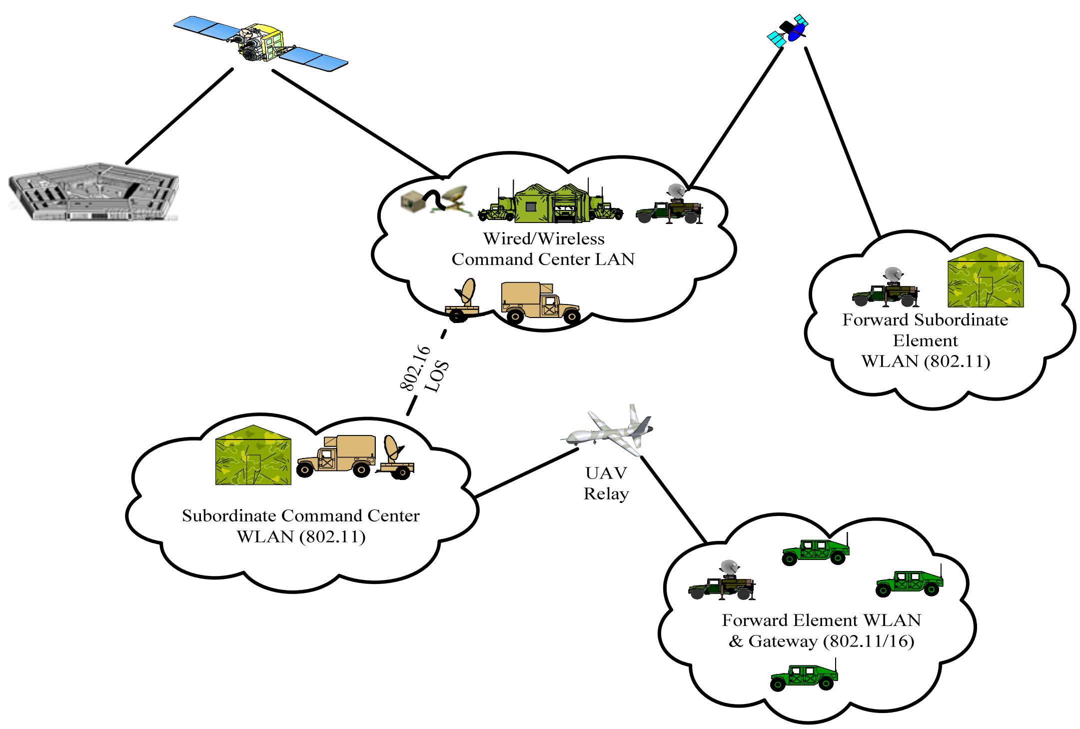

Figure 1. A Vision of Future Converged Battlefield Communication Links

Current VoIP technologies are young and less understood when applied to the wireless domain. Significant wireless VoIP research focus has emerged from the mobile phone community. Industry efforts into VoIP may serve goals that diverge from military specific tactical applications. The prospective savings Department of Defense can achieve through converged system administration, reduced PSTN hardware expenditure, and improved enterprise level efficiency provides a monetary incentive for VoIP research. Economic gains are enhanced by the capabilities set wireless packet-based communication offers to the Coherent Concepts vision.

A. OBJECTIVE

This thesis contains two principal objectives. First, a detailed review of standards for VoIP call signaling and control provides the necessary knowledge to construct a testbed for wireless VoIP implementation. The design provides a scalable architecture to address the need for a flexible VoIP platform for extended research efforts at the Naval Postgraduate School. Operator controlled channel loss replicates the environment packet traffic is most likely to experience during wireless hops. Second, this thesis investigates methods of collecting and presenting voice quality parameters in packet-based networks. Emphasis is placed on non-intrusive, objective voice quality assessment methods that accommodate dynamic testbed topologies. Additionally, speech intelligibility and delay effects are quantified and presented.

B. RELATED WORK

Zhang, Yang, and Quan introduce a simulation framework incorporating wireless links for packet-based voice communications analysis in [4]. System performance and speech quality are examined with an emphasis on applications to the cellular phone market. International Telecommunications Union - Telecommunication Standardization Sector (ITU-T) Recommendations for intrusive network testing are used to extract objective scores via a Perceptive Evaluation of Speech Quality (PESQ) model [5]. Objective scores are compared to the well establish subjective scoring system, also described within ITU-T publications [6].

Zurek, Leffew, and Moreno provide a review of popular objective measurement methods, including PESQ, for VoIP voice quality [7]. A testbed for a packet-based voice network using high compression codecs is described. This research reveals credible correlation between subjective scores and three separate objective assessment techniques for files using G.729 and G.723.1 compression algorithms.

Chemick conducts a fundamental investigation regarding the potential use of voice recognition techniques for voice intelligibility measurement [8]. This work centers on highly compressed digital voice transmissions. Conclusions from the study of voice recognition technologies suggest future work involving the application of commercial software for collection of call intelligibility data. Expansion of this technique is explored in [9] for MATLAB simulated wireless VoIP traffic and popular internet based VoIP services.

Channel simulation using the same hardware available for this thesis is described in a NASA research paper [10] used to validate operation of the Space Communications Protocol Suite Transport Protocol (SCPS-TP). Experiments contained in this publication use the Adtech SX/14 Data Channel Simulator to model ground to satellite conditions for a performance evaluation of transport protocols.

This thesis leverages the lessons of the related material in an effort to extend VoIP quality assessment across a wireless channel. References [4] and [10] were useful guides in recognizing the vision of a wireless VoIP testbed design. Previous work has focused on the implementation of intrusive objective network monitoring techniques. This research effort is based on a non-intrusive approach to objective assessment of voice quality. The combination of lessons from [7] and [9] provide the basis for novel objective measurement methods of call clarity with promising correlation to subjective methods.

C. THESIS ORGANIZATION

This thesis is organized as follows. Chapter II provides a primer on VoIP standards with a focus on the H.323 structure used for this thesis testbed design and experimentation. Chapter III explores the metrics and methods associated with measuring call quality in packet-based communication systems. Chapter IV introduces the testbed designed for this thesis. Chapter V identifies the limitations of the testbed and presents the result of thesis experiments. Chapter VI concludes this study with contributions of this work and suggestions for future expansion and improvement of similar research efforts. Appendix A and B provide a demonstration of step required for data collection and configuring elements of the testbed for experiments, respectively.

THIS PAGE INTENTIONALLY LEFT BLANK

II. INTERNET PROTOCOL TELEPHONY

The evolution of the telephone traverses both analog and digital technologies. The current surge in VoIP interest focuses on a paradigm shift from circuit to packet-based communication. The increased efficiency of packet-based systems drives economic incentive to telecom providers and end users alike. Improvement to a telephony provider network generates cost savings, expanding their ability to serve a growing subscriber population. In contrast, disruptive technologies like VoIP offer more choices for the consumer outside traditional markets. Service providers, such as Skype and Vonage, have thrust Internet-based services to the forefront of modern telecommunications. The acceptance of VoIP within the consumer market will likely depend on a reliable protocol structure that ensures quality and scalability for the future. Goode outlines some of the engineering and standardization challenges to ubiquitous VoIP [11]. This chapter introduces two of the most prevalent standards, Session Initiation Protocol (SIP) and H.323, with an emphasis on H.323 for use in the thesis testbed.

A. SIP

The Internet Engineering Task Force (IETF) introduced the SIP protocol in 1996 as RFC 2543. The most current SIP version is available in RFC 3251 [12]. SIP is often viewed as an approach to IP telephony aligned with web applications or domain name service. SIP only assumes application level signaling duties required to establish a call session. Voice traffic is carried over additional protocols outside of the scope of the RFC 3251. SIP exchanges sequenced messages, similar to Hypertext Transfer Protocol (HTTP), between network elements using a client-server model. A sample call sequence is illustrated in Figure 2. Messages are divided into either request or response categories. Response messages also split into a numbered class system. Examples of the request and response message format are shown in Table 1. This fairly simple structure has made SIP an attractive alternative to the more complex H.323.

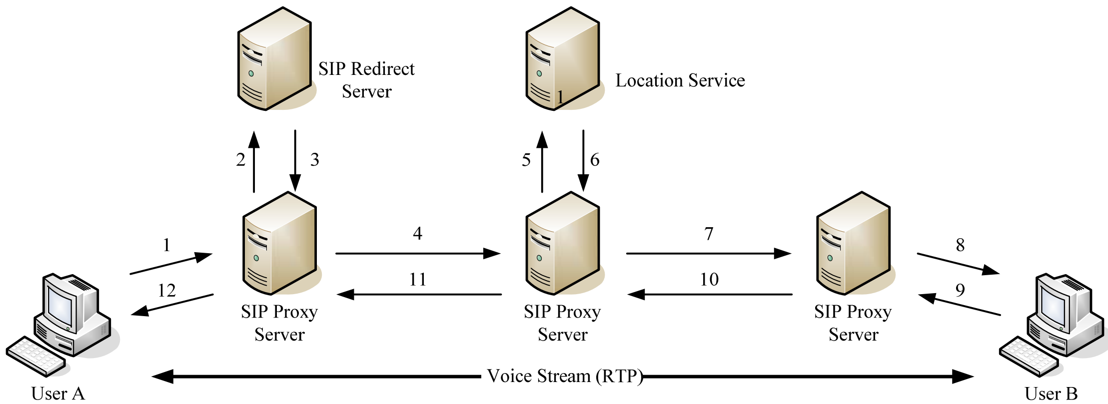

Figure 2. SIP Call Sequence: User A initiates a voice call to User B

SIP Request	Purpose
INVITE	Invite a user to a call
ACK	Acknowledge
OPTIONS	Get server capabilities
BYE	Close or deny call
CANCEL	Terminate action
REGISTER	User Location Report
INFO	Mid session signal

Response Classes	Purpose
1XX	Informational
2XX	Successful
3XX	Redirect
4XX	Client Error
5XX	Server Error
6XX	Global Failures

Table 1. SIP Request and Response Formats

As with any young IETF protocol, there are still issues ripe for debate and improvement through the RFC process. SIP has faced some PSTN interoperability challenges during the first decade of use [13]. Such limitations have, in part, led to greater market penetration of H.323 based hardware. Undoubtedly, the continued evolution of SIP will provide some of the most serious competition among VoIP standards.

B. H.323

The oldest and most prevalent VoIP protocol in use is ITU-T Recommendation H.323. Its initial release took place in 1995 under the name, "Visual Telephone Systems and Equipment for Local Area Networks Which Provide a Non-guaranteed Quality of Service." H.323 version 2, changed the name to "Packet-based Multimedia Communications Systems." Version 6, released in 2006, is the most current update of the H.323 standard [14].

When the ITU-T set out to address the growing demand for a protocol addressing transmissions across packet networks, they turned to the existing H.32X family of protocols. This collection of ITU-T Recommendations governs multimedia transfer across disparate networks. Figure 3 shows the interrelationship of H.32X series protocols. One product of this lineage has been an intense focus on interoperability with diverse worldwide telecommunications systems. Protocol design challenges are magnified by the appetite for more powerful combined services (e.g., video conferencing). In this light, VoIP has merely surfaced as the most visible application of choice. The remaining sections of this chapter explore the components and control structures required for proper VoIP operation in a network using H.323.

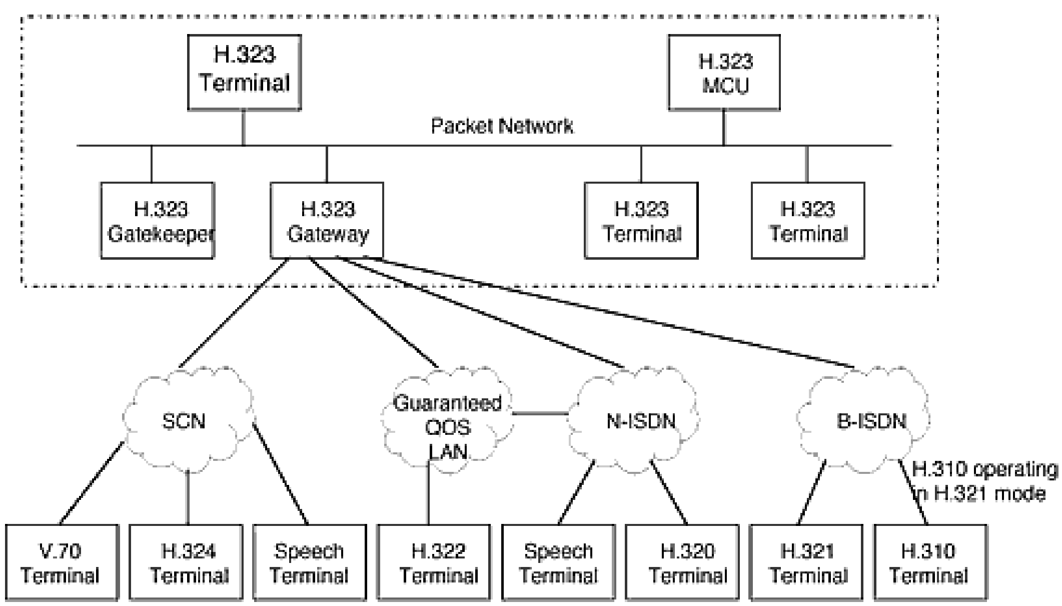

Figure 3. ITU-T Recommendation H.32X Family (from [15])

C. H.323 COMPONENTS

The scale and structure of any H.323 VoIP network can vary widely based on the needs of the users it is designed to service. Typical large scale fielding of voice service requires several administrative areas subdivided into subordinate elements. These divisions often take place along geographic or management boundaries (e.g., cities and facilities). The basic building blocks of these networks are VoIP zones. Each zone contains a variable mix of the four fundamental H.323 components. Logically, these are individual components. Some hardware (e.g., Cisco routers) can combine logical duties within a single physical device [14]. The top of Figure 3 shows a sample VoIP zone.

1. Terminals

Terminals act as the human interface for a real time, full duplex multimedia exchange. H.323 requires all standard compliant terminals to offer audio session support. Video and data capabilities are an optional extension to basic voice service. Terminals can be PCs or stand alone devices. H.323 terminals are compatible with terminals from the full H.32X family of protocols.

2. Gateways

In VoIP structures, there are three general call architectures describing connections between terminal types, IP to IP, non-IP to IP, and non-IP to non-IP. A gateway allows H.323 terminals to share multimedia with dissimilar networks.

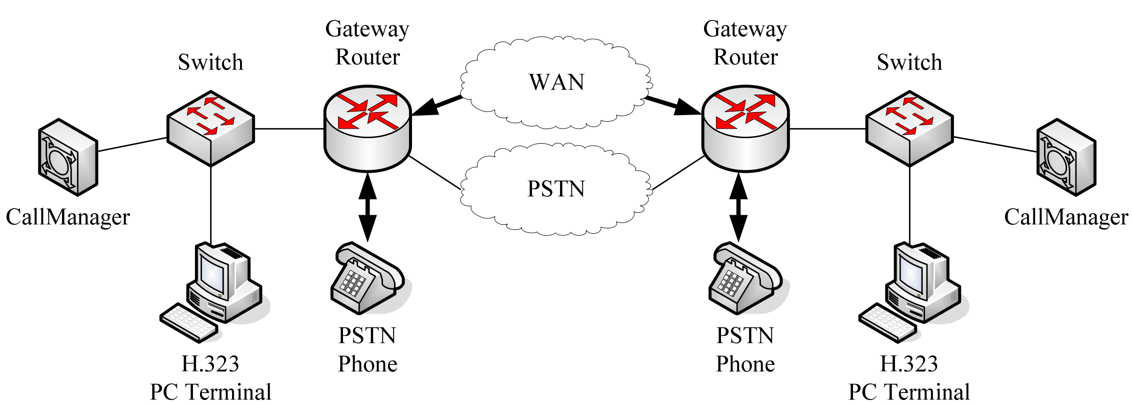

Figure 4. H.323 Gateways with PSTN Bypass

Figure 4 shows gateways used for voice stream translation and toll bypass of normal PSTN service. This format is common in organizations that want to reduce flow across high cost connections. PSTN or alternate trunks are often maintained for redundancy. Call connections are possible for all combinations of the associated terminals in the illustration. There is no defined limit to the number of gateways within a VoIP zone.

3. Gatekeepers

Gatekeepers perform tasks, such as admission control, address translation, billing, and gateway management. As the scale of VoIP zones increases there are often competing interests for limited resources on the converged packet network. Gatekeepers have the ability to control bandwidth allocation to registered terminals. Additional functions include directory and call control assistance. Gatekeepers are an optional component within the H.323 standard. When used, only one gatekeeper may reside per VoIP zone.

4. Multipoint Control Units

Multipoint Control Units (MCU) are composed of a Multipoint Controller (MC) and an optional number of Multipoint Processors (MP). Combined, these units conduct call control for conferences of three or more multimedia endpoints. The MCU carries out the capability exchange and selection of communication mode for conference sessions. MCUs may have the ability to convert between different media formats (audio, video, and data), and bit rates among terminal devices.

D. H.323 SIGNALING AND CONTROL

Call signaling and control define the logical measures required to setup, maintain, and teardown a multimedia session. H.323 enlists a collection of protocols, shown in Figure 5, to accomplish the mixture of tasks necessary for managing communication links. The TCP/IP suite provides a solid foundation for reliable and best effort transport of H.323 related messaging. This section will explore those signal and control structures critical to VoIP applications. An introduction to Real-time Transport Protocol (RTP) is included.

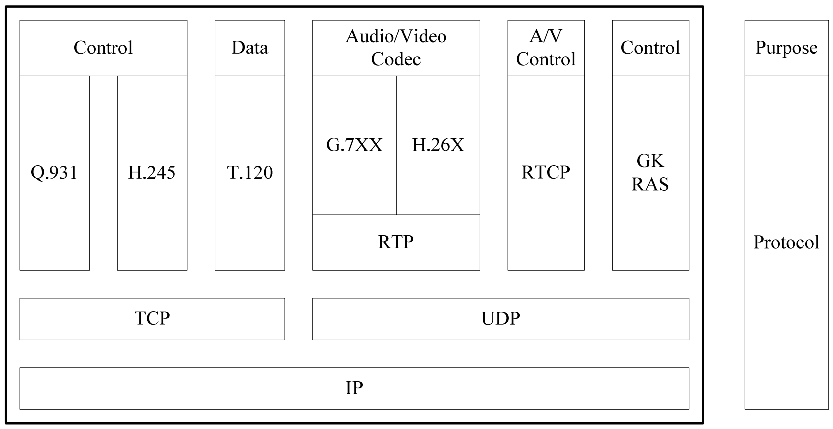

Figure 5. H.323 Protocol Relationships

1. H.225.0 Registration, Admission, and Status (RAS)

Gatekeeper components employ the RAS to convey registration, admissions, bandwidth change, and status messages. Exchanges take place across an unreliable channel via User Datagram Protocol (UDP) subject to timeout and retransmission. During the termination phase of a call sequence, this channel handles disengagement of registered endpoints from the assigned gatekeeper. Detailed review of gatekeeper messaging is available from [14] and [16].

2. H.225.0 Call Signaling

The call setup process shifts from the RAS channel to a reliable TCP connection for endpoint signaling. The H.225.0 call signaling channel is designed to manage concurrent call requests. All messages conform to the Q.931 Integrated Services Digital Network (ISDN) control format [17]. Networks equipped with a gatekeeper select one of two options for H.225.0 message routing. In the absence of a gatekeeper, signaling passes between endpoints.

a. Direct Endpoint Signaling

When direct endpoint signaling is used, the source component starts the process by sending an admission request to the gatekeeper on the RAS channel. The gatekeeper confirms or rejects the request according to configured management parameters via the same RAS channel. Confirmation results in a setup message transmission from the source endpoint directly to the target endpoint. After a final RAS exchange the receiver endpoint responds with a connect message.

This signaling structure allows the gatekeeper to manage bandwidth and accounting while distributing some of the processing action among endpoints. Call volume and duration data can be stored from the RAS and disengage messaging that bracket each session. Figure 6 illustrates a direct endpoint signaling exchange. This model can also be extended to more complex architectures using multiple gatekeepers. Extensive discussion of scaled network design, with an emphasis on call control, can be found in [18]. Networks void of gatekeepers use direct endpoint signaling without a RAS exchange.

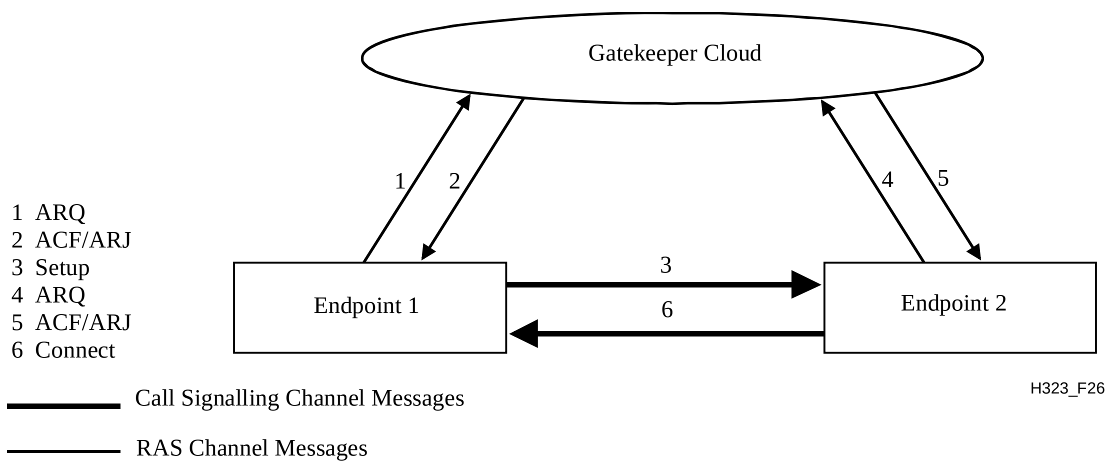

Figure 6. Direct Endpoint Signaling (from [14])

b. Gatekeeper Routed Signaling

Gatekeeper routed call signaling is an alternative call control format to direct endpoint signaling. This form of routing forces all signaling traffic flow along a strict path through a gatekeeper. Consequently, greater overall message volume is required to establish a communication session using gatekeeper router signaling. Figure 7 illustrates a direct endpoint signaling exchange. Cisco IOS does not support this form of routing within gatekeeper components [19].

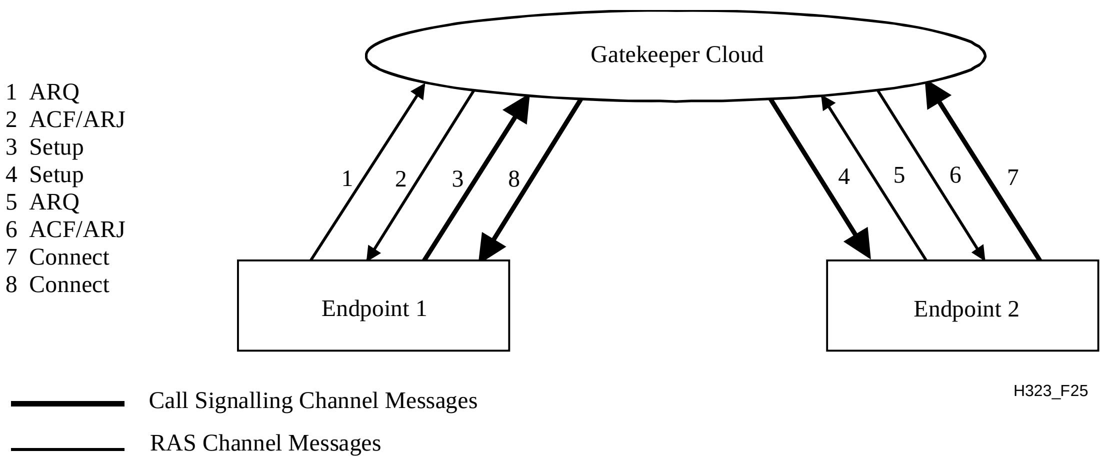

Figure 7. Gatekeeper Routed Signaling (from [14])

3. H.245 Call Control

After the initial signaling for a multimedia session is complete, call control messaging establishes additional coordination between endpoints prior to the start of multimedia transmission. H.323 conducts call control using the H.245 protocol detailed in [20]. The H.245 call control channel is governed by the same direct or gatekeeper enabled path options that manage H.225.0 flow. This thesis will focus on the direct call control model.

H.245 messages can be grouped into four categories: request, response, command, and indication. Endpoints use H.245 to elect a master multipoint controller, exchange Terminal Capability Set (TCS), and agree on communications procedures supported by all parties. H.245 is also responsible for establishing a logical channel for multimedia transfer. This logical channel remains open for the duration of a call session. Additional flow control and general purpose commands complete the basic H.245 functions.

4. Audio Codecs

One key portion of the H.245 TCS exchange for a VoIP session involves the audio codec established for the logical channel voice stream. Codecs convert and compress the voice signal into a scaled bit stream for transport, but the application of a codec is an isolated segment of the larger signal processing path. Figure 8 illustrates the general signal flow.

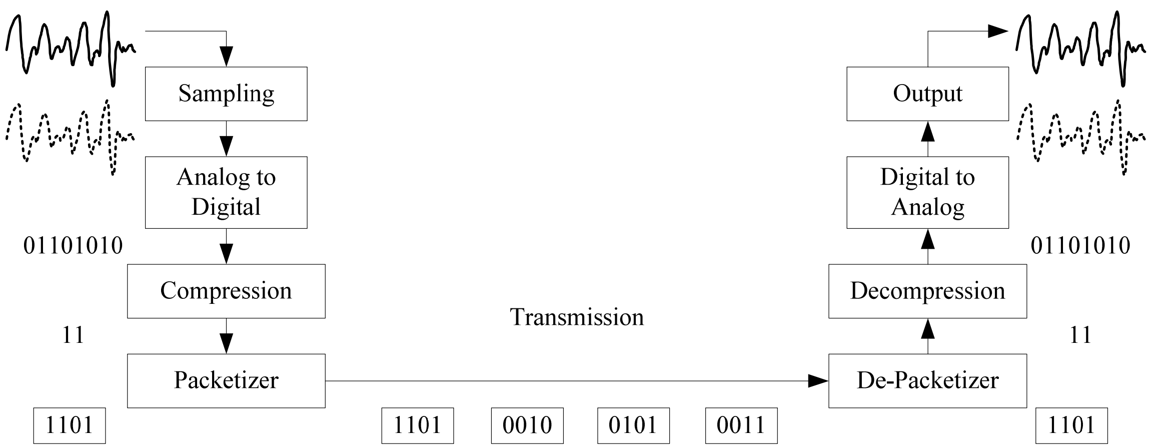

Figure 8. Signal Processing Steps

The voice signal arriving at a terminal microphone is typically sampled at 8000 Hz, preserving spectral content up to 4000 Hz and below for processing and reconstruction [21]. Samples are transformed into a digital representation of the original waveform according to the codec specification and compression algorithm. The sample rate, sample size, and compression ratio determine the bit rate of a codec. As the packets are prepared for transmission, each codec provides a different size block of data for the voice payload. Table 2 contains a comparison of popular codecs maintained under the ITU-T G.7XX family of recommendations [22, 23]. All H.323 terminals are required to support G.711.

Codec		Voice Block Size (bytes)	Compression Ratio	Bit Rate (kbps)
G.711	PCM	80	1:1	64.0
G.723.1	MP-MLQ	240	10:1	6.3
G.723.1	MP-ACELP	240	12:1	5.3
G.726	AD-PCM	80	2:1	32.0
G.728	LD-CELP	80	4:1	16.0
G.729A	CS-ACELP	80	8:1	8.0

Table 2. Codec Comparison (after [24])

5. Real-Time Transport Protocol (RTP)

RTP is an IETF protocol [25] designed to support the real-time transfer of data between two or more members of a multimedia session. Riding above the UDP transport layer, RTP focuses on providing timely media delivery rather than reliable services to session participants. VoIP calls in an H.323 system pass packetized bit streams from the codec down the RTP-UDP-IP stack. A typical link level packet format is shown in Figure 9.

x bytes	20 bytes	8 bytes	12 bytes	x bytes
Link Header	IP Header	UDP Header	RTP Header	Voice Payload

Figure 9. VoIP Packet Structure

RTP header values include data source, timestamp, sequence, and payload identification fields to assist in the recovery of media packet data. Sequence and time information facilitate endpoint activities to defeat negative network effects to packet delivery. Buffers allow sequence and time data to assist during reconstruction of original packet order and a reduction in delay variation for final transmission. RTP header values also facilitate network statistical analysis by tracking the distribution and rate of packet loss. RTP does not provide any form of error detection or control. Figure 10 provides a detailed view of the common VoIP header fields.

RTP Control Protocol (RTCP) is a companion protocol defined within RFC 3550. RTCP manages quality of service, identification, session scaling, and session control of the RTP stream [26]. RTCP packets are issued periodically, using a separate port number, to session members in a multicast fashion.

RTP

<table>
<tr><td>V</td><td>P</td><td>X</td><td>CC</td><td>M</td><td>Payload Type</td><td>Sequence Number</td></tr>
<tr><td colspan="7">Timestamp</td></tr>
<tr><td colspan="7">Synchronization Control Source</td></tr>
</table>

UDP

<table>
<tr><td>Source Port</td><td>Destination Port</td></tr>
<tr><td>Length</td><td>Checksum</td></tr>
</table>

IP

<table>
<tr><td>V</td><td>HL</td><td>TOS</td><td colspan="2">Total Length</td></tr>
<tr><td colspan="3">Identification</td><td>Flags</td><td>Fragment Offset</td></tr>
<tr><td colspan="2">TTL</td><td>Protocol</td><td colspan="2">Header Checksum</td></tr>
<tr><td colspan="5">Source IP Address</td></tr>
<tr><td colspan="5">Destination IP Address</td></tr>
<tr><td colspan="5">Options</td></tr>
</table>

Figure 10. RTP-UDP-IP Headers

E. H.323 VOIP CALL SEQUENCE

Signaling tasks in a VoIP call sequence are divided into five phases [14]. This section focuses on actions carried out during the signaling phases related to a VoIP call sequence for networks void of any gatekeeper component.

1. Call Setup

Call setup, the first phase of the call sequence, proceeds according to the configuration of components on each end of a potential multimedia exchange. In the absence of a gatekeeper, endpoints conduct direct signaling and bypass the need for bandwidth reservation requests. The lack of endpoint synchronization during this phase introduces the risk of simultaneous setup requests. To handle the potential for concurrent requests, endpoints provide a busy response to incoming call requests while waiting for replies from their own setup messages. Endpoints expect a response within four seconds of a successful setup message transmission. Figure 11 shows the call setup message sequence with direct signaling.

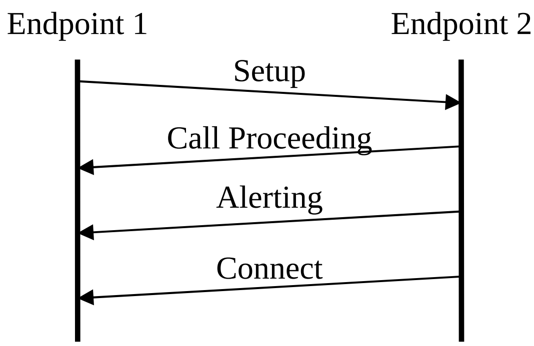

Figure 11. Direct Endpoint Routing Call Setup Message Exchange

2. Initial Communications and Capability Exchange

After endpoints exchange call setup information, they establish a direct H.245 channel. TCS information starts the H.245 message flow through the control channel. Following confirmation from both sides, via TCS Ack messages, the codec is selected for VoIP service. If any interruption occurs during the TCS exchange, the control process stops and reinitiates a new TCS message. Endpoints that receive a TCS halt active communication until they can respond and negotiate the required channel controls. Following TCS messaging, the endpoints conduct a Master/Slave Determination (MSD) to elect the active MC device for any conference call events. All message exchanges are permitted up to three total transmissions before a communication failure is tagged within this phase. Retransmission failures result in a shift from the capability exchange phase to call termination. Figure 12 depicts a successful direct endpoint TCS and MSD exchange.

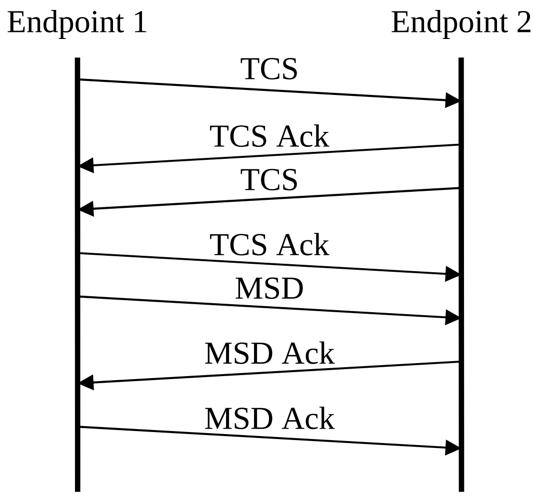

Figure 12. Capability Exchange and Master Slave Determination Sequence

3. Establishment of Audiovisual Communication

The third phase of the call sequence opens a logical channel configured for the type of multimedia transfer among the select number of endpoints. Audio specific applications, like VoIP, ride on the unreliable RTP-UDP-IP stack. The remaining actions available within this phase are associated to multipoint audio conferencing or logical channel control for video transfer. Figure 13 illustrates the message exchange used to open a logical channel for the typical two-party VoIP applications.

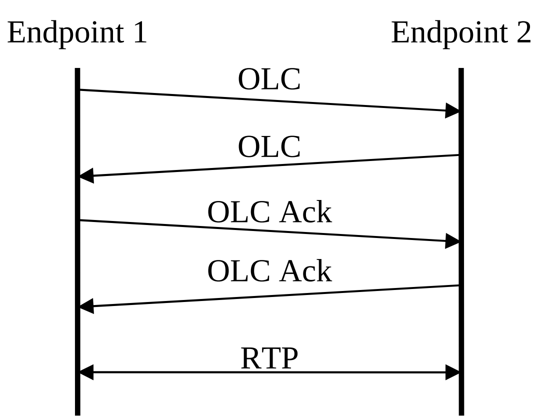

Figure 13. Control Message Exchange to Open Logical Channel

Alternate audio oriented options to the message flow include media stream address distribution, conference matching to RTP streams, and communications mode command procedures. MCU components conduct address assignment for conference endpoints. The MC element of the MCU determines the unicast or multicast structure of conference sessions. The MC can direct the open and close of logical channels to achieve the desired centralized or decentralized control format of the conference.

4. Call Services

Once the VoIP RTP stream has been established, a group of H.245 commands provide additional services during the active call period. Variable rate codecs and bandwidth controlled networks have the ability to apply bandwidth changes to a call in progress. These channel modifications are carried out by closing the original logical channel, opening a new updated logical channel, and seamlessly transferring user traffic to the new connection.

Phase four of the call sequence also allows ad hoc conference expansion. Figure 14 shows a new user (Endpoint 3) negotiating admittance to an active call. The joining endpoint transmits a setup request including user identity, target Conference Identifier (CID), and intentions. Message sequencing for call services depends heavily on network component architecture and the active MC selected from previous signaling phases. Detailed message flow for complex topologies can be found in [14].

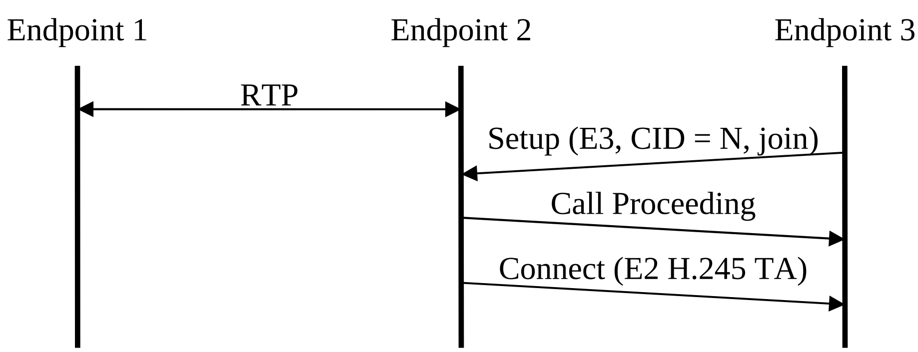

Figure 14. New Endpoint Admittance to Ad Hoc Conference

Additional supplementary services are offered to H.323 endpoints according to network configuration. These extensions are defined within ITU-T H.450.X series of recommendations [27]. Services include common telephony features, such as call transfer, hold, diversion, and caller ID.

5. Call Termination

The conclusion of the call sequence carries out the termination of logical channels. Any endpoint or immediate call signaling entity can initiate the termination phase. Figure 15 shows an example of endpoint directed call termination. The end session command halts all media transmission prior to closing logical channels associated to the session. In the event of control channel failure during an active VoIP call, H.323 prevents immediate call termination. If a means to re-establish failed H.225.0 or H.245 signaling exists, the VoIP application will continue during a recovery effort. The absence of any means to recover call control will initiate the termination sequence.

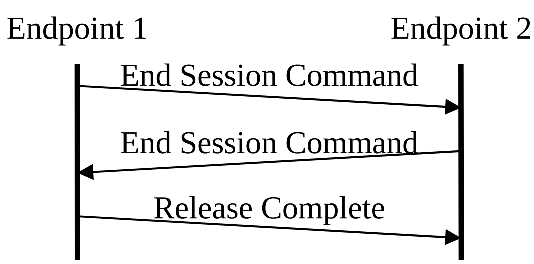

Figure 15. Endpoint Directed Call Termination Control Messages

F. SUMMARY

VoIP is an emerging multimedia application poised to revolutionize voice communications. This chapter introduced the prominent VoIP enabling protocols used today. H.323 components, signaling, and call sequence were presented with a focus on direct routing implementation. The focus on VoIP network design will now shift to the metrics and methods recommended in support of VoIP performance analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

III. VOIP PERFORMANCE

While traditional telephony enjoys a long history of performance evaluation and testing, VoIP is fairly new and presents unique challenges. This chapter introduces the metrics and techniques used to assess voice quality in packet networks. VoIP performance testing schemes and predictive electronic tools are studied from the perspective of cost, accuracy, and scalability. Two approaches to voice recognition are presented. These elements combine to form a foundation for the evaluation of thesis testbed data.

A. VOICE QUALITY METRICS

Before measurement and analysis of any network can take place, an observer must identify proper metrics for data collection. This section examines voice quality as a function of delay, echo, and clarity [28]. Figure 16 illustrates the conceptual relationship of these variables to the human perception of speech quality. An ideal network resides at the plot origin, where data delivery is instantaneous with no echo and perfect clarity. The point representing voice quality moves away from the origin as realistic impairment factors are considered.

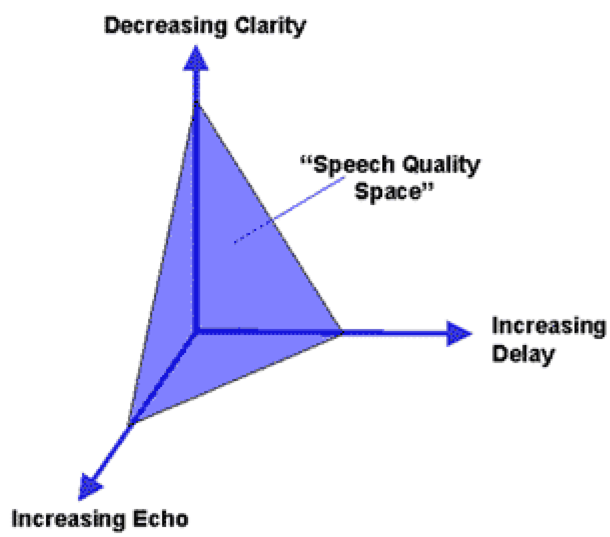

Figure 16. Relationship of Delay, Echo, and Clarity to Voice Quality (from [28])

1. Delay

Delay is defined as the amount of time required for a signal to traverse a network. Isolated forms of delay can be categorized by the fixed or variable contributions they provide to the cumulative end-to-end delay of a network. Increasing amounts of delay tend to impose negative effects on call quality by forcing a half-duplex style conversation onto users. Recommended values of delay for voice applications are established in [29]. Figure 17 shows estimated user satisfaction for different delay values. The plot uses a predictive modeling tool discussed later in this chapter.

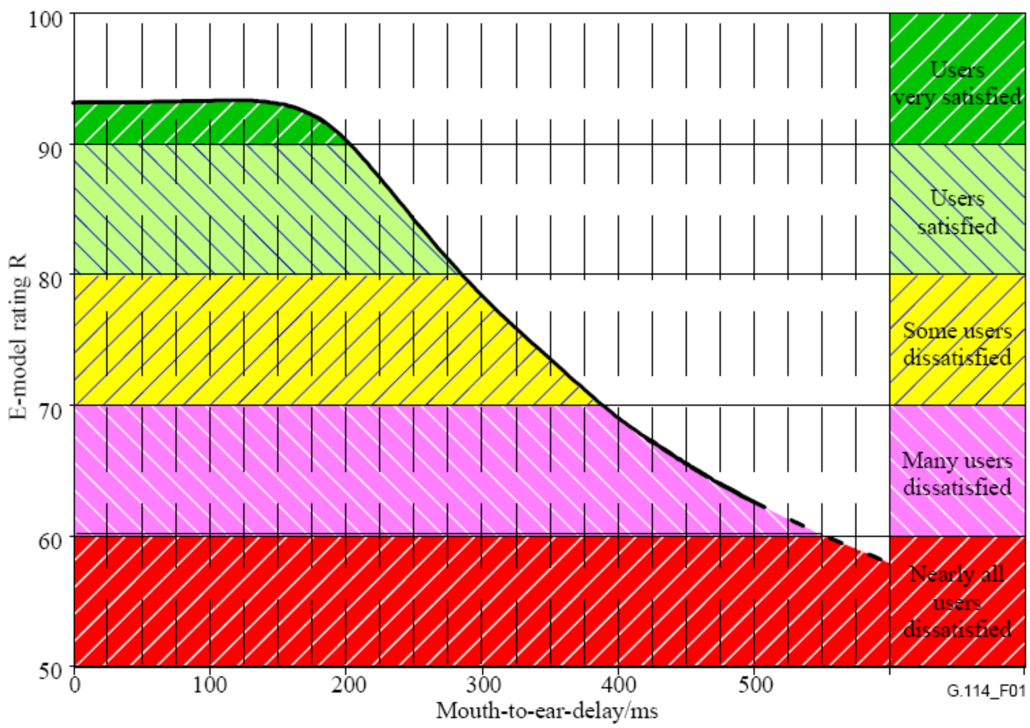

Figure 17. Effect of Delay on User Satisfaction Estimated by E-model (from [29])

Cisco Systems has summarized the critical sources of delay for packet networks in [30]. Fixed delay can be attributed to several actions necessary to prepare and transport packets. Codecs require a predictable number of clock cycles to read, compress, and de-compress voice data. For example, the typical processing delay for G.729 amounts to 18 ms. More fixed time is lost as the payload of each packet is filled with data, known as packetization delay. Next, serialization delay accounts for the transmission time required for frames to enter the network. Finally, propagation delay between endpoints will vary according to link distance and the physical channel. In long distance networks, signal propagation accounts for a majority of the fixed delay.

Variable sources of delay provide a random element to the end-to-end cumulative value. Propagation distance is only assumed to be a fixed value for individual packets. Random delay variation, called jitter, surfaces as packets take different paths though the network. Packets also face non-uniform queuing delay while they compete for access to the physical medium. The length of queues can change drastically based on local traffic loads and wide area network factors. To reduce the impact of jitter, additional buffers are employed to ensure a relatively constant stream of voice packets is available to the receiver. Modern jitter buffers contribute a variable delay since their length adapts to the statistics of arriving packet streams [30].

2. Echo

Echo occurs in telephony applications when a talker's voice returns to their own receiver. This form of impairment is most prevalent in VoIP networks connected to the PSTN. Echoes are primarily generated by an impedance mismatch within electrical junctions. Unbalanced circuits are most common in connections where four-wire or digital transmission lines are converted into separate two-wire transmit and receive segments. Traffic on the listening side of the network leaks from the receive line into the transmission path at these junctions [21]. A secondary impairment, called acoustic echo, is generated when output from a terminal speaker couples to the microphone [30].

The impact of echo can be reduced by deploying echo cancellers at different locations within the network. Cancellers are devices that monitor voice activity and mathematically model the probable echo. Impairment effects are removed by combining regular voice traffic with a negative version of the modeled echo. Contemporary VoIP terminals incorporate echo canceling algorithms that adapt and converge to a corrective model for the current voice session [30].

Delay and attenuation of echo along the transmission path helps determine the level of impairment encountered during a conversation. Figure 18 identifies acceptable echo characteristics according to one-way transmission delay and Talker Echo Loudness Rating (TELR). TELR is a measure of attenuation the echo encounters along the round trip path through a network. In general, people tolerate the loudness rating of an echo less as delay increases. Methods for calculating TELR are defined in [31].

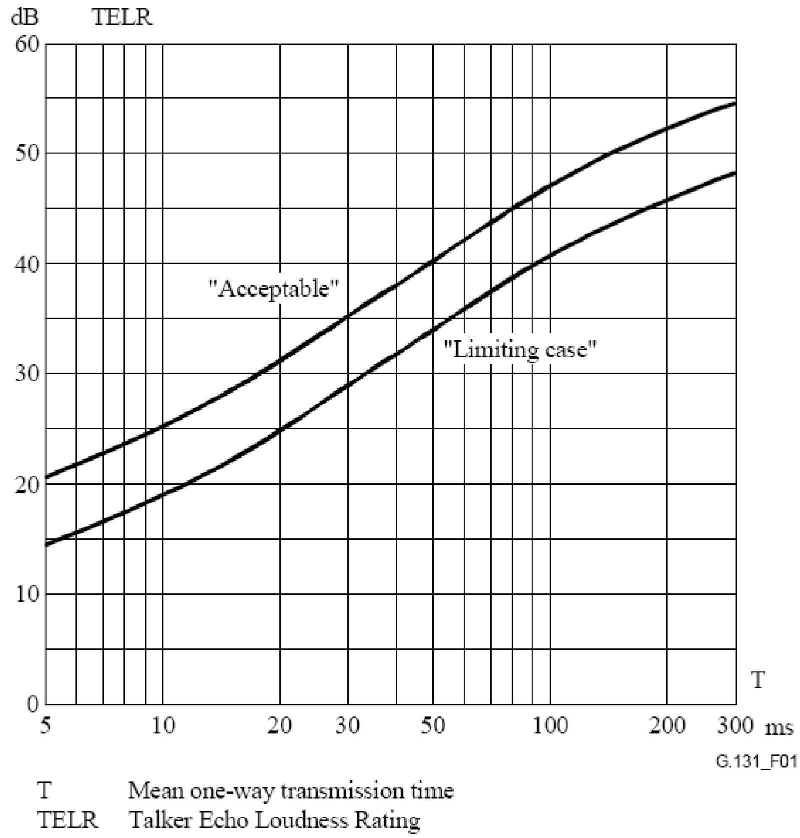

Figure 18. Listener Tolerance of Talker Echo (from [31])

3. Clarity

Clarity has the most expansive and subjective interpretation among the voice quality metrics. The Internet Engineering Consortium defines clarity as the perceptual fidelity, clearness, and the non-distorted nature of a particular voice signal [28]. Intelligibility of speech is often implied when describing clarity, but comprehension of spoken words does not always equate to a clear voice signal free of distortion. It is possible to extract content from a sentence of poorly reproduced speech. This apparent contradiction in defining clarity reveals the challenges that emerge when defining the complex subjective nature of human verbal communication. The interaction of clarity and intelligibility are managed differently by each assessment approach. This section will introduce key factors that impact clarity and exhibit a potential to degrade the comprehension of verbal signal content.

Noise is a diverse and persistent source of impairment to voice clarity. In general, noise will manifest in the form of environmental factors, analog circuitry contributions, and bit errors. Background noise entering a phone, or the receiver's listening environment, can be regulated for testing events and daily use. The factors of greater interest are those which cannot be readily altered by a user, such as bit errors attributed to a wireless channel. Noise corrupts and distorts the speech reproduced at VoIP terminals [28].

Packet loss robs the listener of entire speech blocks, degrading the perception of voice clarity. Loss on this scale is often a function of network congestion. When traffic volume reaches an unsustainable level buffers overflow, and packets that cannot be queued for transmission are dropped. Time sensitive applications like VoIP also suffer packet loss when delay in packet arrival exceeds the bounds of the de-jitter buffer. Any perceived benefit in a lengthy de-jitter buffer must be balanced against the contributions in end-to-end delay [28].

Codecs assist in the management of network bandwidth at the cost of delay and clarity. Every increase in codec compression ratio and complexity results in greater processing delay. Clarity also declines when increased compression is used. As fewer data bits are used to describe voice content, an algorithm's ability to reconstruct the detailed perceptive elements of speech declines [28].

B. VOICE QUALITY ASSESSMENT AND PREDICTION

Voice quality has been the subject of intense study over the past century. Telecommunications providers view voice quality perception as the key economic driver in the industry. Understandably, there are a variety of assessment tools and methodologies that have evolved with the modern telephony applications. Within the last decade, most popular voice quality standards have posted updates or extensions to address VoIP specific concerns. This chapter provides an introduction to current assessment techniques with a focus on cost, accuracy, and scalability to a VoIP testbed.

1. Subjective Assessment of Voice Quality

The oldest and most fundamental of the assessment techniques is the ITU-T recommendation on methods for subjective determination of transmission quality [6]. This document provides testing format and grading guidance for telephony experiments attempting to capture direct human perceptions of performance. Typical testing includes a five-level grading scale for the categories of listening-quality, listening-effort, and loudness-preference. Each category is assigned a numerical score according to the description in Table 3. These grades form a subjective measurement scale known as the Mean Opinion Score (MOS). This thesis will focus on results related to MOS for the listening-quality scale.

MOS	Listening-Quality Scale	Listening-Effort Scale	Loudness-Preference Scale
5	Excellent	Complete relaxation; no effort required	Much louder than preferred
4	Good	Attention required; no appreciable effort	Louder than preferred
3	Fair	Moderate effort required	Preferred
2	Poor	Considerable effort required	Quieter than preferred
1	Bad	No meaning understood	Much quieter than preferred

Table 3. MOS Grading Scale and Description

Large scale subjective testing, polling several thousand subjects, is prized for capturing the intangible elements of psychology and mood. MOS represents the benchmark all remaining techniques seek to replicate.

2. Objective Assessment of Voice Quality

Unfortunately, subjective MOS is rarely scalable or practical for the fluid collection of data in a testbed. The time and cost associated with human subjects are often prohibitive. These limitations have served as industry drivers for accurate objective voice quality assessment techniques.

In response to testing needs, the ITU-T published recommendations P.862 Perceptive Evaluation of Speech Quality (PESQ), and P.563. These standards provide computer based assessment models capable of mapping objective assessment data to a MOS-LQO (Listening Quality Objective) mirroring subjective scores. Methods are distinguished by the manner in which they collect voice information for model processing. Figure 19 compares the intrusive PESQ (P.862) testing schematic with the non-intrusive P.563 format. Objective assessment methods have shown the ability to map MOS-LQO results with an error less than 0.25 MOS (± 0.25 on a 5-point scale) for 72.3% of validation test conditions [5, 32].

This thesis utilizes a pre-standard, objective, single-ended model related to P.563 for baseline voice assessment. Non-intrusive methods still exhibit limitations in their ability to assess channel delay characteristics. The next section explores an ITU tool for predictive network modeling that addresses variable delay considerations.

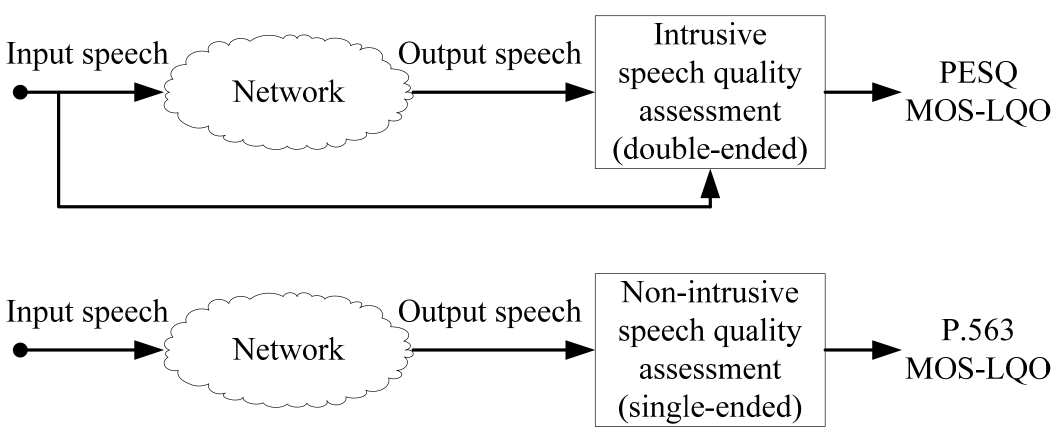

Figure 19. Comparison of Intrusive and Non-intrusive Assessment Setup

3. Predictive Voice Quality Modeling

Each of the preceding assessment techniques was designed to test voice quality within an established network. Results from objective VoIP tests rarely translate into forward looking design recommendations. This section presents a computational tool, known as the E-model, intended to aid engineers in transmission and network planning [33].

The E-model is a predictive mathematic representation of network impairments defined by component selection and the physical channel. Psychological effects of each impairment factor are considered additive in nature. The cumulative representation of elements is captured in the transmission rating factor, R, given by

$$R = SNR_O - I_s - I_d - I_{e,eff} + A \tag{3.1}$$

where:

SNR_O	signal-to-noise ratio,
I_s	impairments simultaneous to the signal,
I_d	impairments from delay,
$I_{e,eff}$	packet loss, impairments from equipment (e.g., codec), and
A	advantage factor (e.g., elevated tolerance for mobility convenience).

This thesis uses the E-model to explore the impact of link delay on R value. The delay impairment factor, I_d, can be isolated and divided into three factors

$$I_d = I_{dte} + I_{dle} + I_{dd} \tag{3.2}$$

where I_{dte} represents impairments from talker echo, I_{dle} represents impairments from listener echo, and I_{dd} represents impairments excessive absolute delay. Current hardware embedded echo cancellation results in the domination of I_d by the I_{dd} term.

Specific values of I_{dd} can be calculated using

$$\text{For } T_a \le 100 \text{ ms:} \qquad I_{dd} = 0$$
$$\text{For } T_a > 100 \text{ ms:} \qquad I_{dd} = 25\left\{\left(1+X^6\right)^{\frac{1}{6}} - 3\left(1+\left[\frac{X}{3}\right]^6\right)^{\frac{1}{6}} + 2\right\} \tag{3.3}$$

with

$$X = \frac{\log_{10}\left(\frac{T_a}{100}\right)}{\log_{10}(2)} \tag{3.4}$$

where T_a is the absolute delay [33]. After impairments are incorporated into the transmission rating factor, conversion to an estimated subjective score helps predict user satisfaction. The R value to MOS conversational quality estimate (MOS_{CQE}) is calculated as follows:

$$\text{For } R < 0: \qquad MOS_{CQE} = 1$$
$$\text{For } 0 < R < 100: \quad MOS_{CQE} = 1 + 0.035R + R(R-60)(100-R)7 \cdot 10^{-6} \tag{3.5}$$
$$\text{For } R > 100: \qquad MOS_{CQE} = 4.5$$

where the range of $6.5 < R < 100$ bounds the valid range for the equation to calculate an R value from MOS_{CQE}. Figure 20 illustrates the mapping of R value to MOS_{CQE} [33].

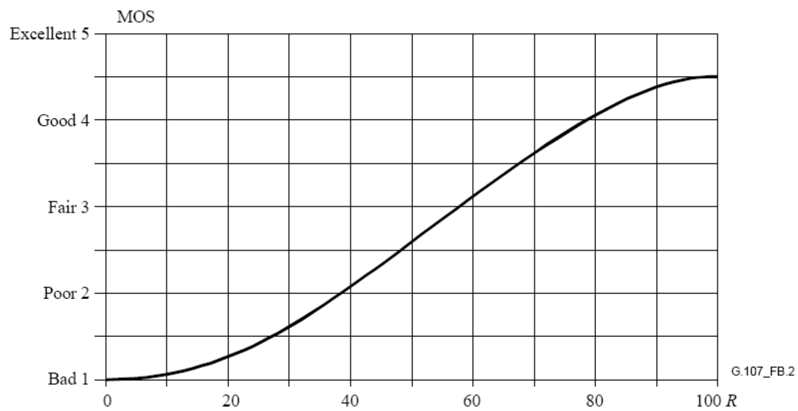

Figure 20. R Value to MOS_{CQE} Conversion (from [33])

C. VOICE RECOGNITION

Voice recognition is a technology that allows machines to artificially comprehend and act upon received voice signals. Acceptable performance in early systems was limited by vocabulary size, speaker constraints, and specific conversational tasking (e.g., dialing a telephone number). Modern systems aim to handle conditions more aligned with natural human conversation. Current technologies devoted to recognition use isolated word recognition (IWR) or continuous speech recognition (CSR) depending on user needs [34]. This section introduces common processing techniques associated with IWR and CSR.

1. Dynamic Time Warping

Recognition of speech signals is complicated by the random temporal attributes of speaker behavior. A person uttering a word or syllable produces subtle variations for each realization of a measured speech element. First generation voice recognition algorithms resolve temporal changes with a template matching scheme, called Dynamic Time Warping (DTW) [34].

DTW applies a trained reference template to an observed voice sample element (e.g., a single word or phoneme). A mathematic tool, dynamic programming, analyzes the files for optimal decision matching. By temporally stretching or compressing the reference file, it can be "warped" in time to provide symmetry with observations. Practical applications require well defined speech element boundaries for successful DTW application. DTW-based recognition typically focuses on IWR where speakers are confined to cooperative situations with limited vocabulary. CSR is possible using DTW, but template length and computational expense prohibit suitable scalability for commercial applications [34].

2. Hidden Markov Model

DTW templates fail to address the inherent variability associated with a non-ideal speaker in CSR. A human physiologic structure produces different variations of a discrete sound based on inter word relationships. Transitions within a language are defined by the lexical and syntactic rules that govern linguistic structure. Contemporary voice recognition accounts for speaker variability by modeling sound production as a stochastic process. The most prevalent method for CSR is the Hidden Markov Model (HMM). This form of speech processing takes place in two phases, training and recognition [34].

During the training phase, an HMM examines a reference file and stores statistical characteristics of spoken units (e.g., sentences, words, and phonemes). Analysis reveals mathematical features of the isolated speech units, states, and the relationships extending to neighboring states. Complex CSR requires feature resolution to the sub-word level. English, for example, contains approximately forty-two distinct sounds for word construction. The HMM can exploit statistical aspects of both acoustic production and language structure. Figure 21 illustrates the finite compilation of state associations that define a given HMM. Numbered states represent the variable form of word units and grammatical organization.

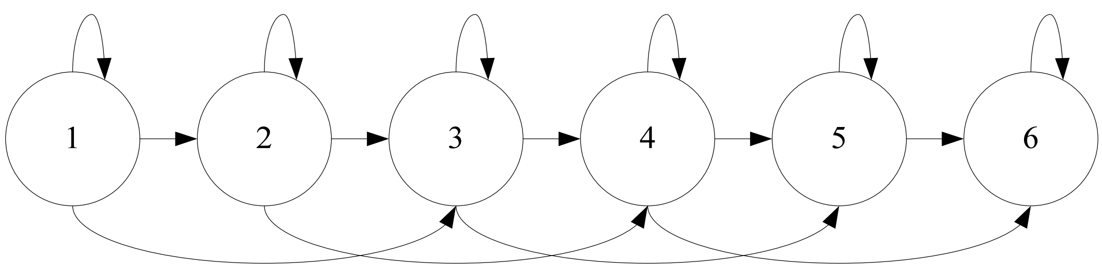

Figure 21. Six State HMM

The recognition phase treats the HMM as a finite state machine. Sampled voice streams supply the model with observations. Words are recognized by comparing the trained HMM to the incoming stream. One stored model provides the highest likelihood of generating the observed string, and represents the designated match. So far, HMM applications have demonstrated CSR capabilities superior to DTW [34]. Dragon NaturallySpeaking is a HMM-based voice recognition tool used in this thesis.

D. SUMMARY

This chapter introduced the voice quality metrics of delay, echo, and clarity. Factors that contribute to the behavior of each metric were explored in relation to a VoIP network. A primer on ITU-T recommended methods for assessing and predicting voice quality was provided. Conceptual approaches and techniques for voice recognition were briefly presented.

IV. TESTBED DESIGN

This thesis develops a testbed designed to carry packet-based multimedia communications using the H.323 standard. Cisco Systems Unified Voice products are deployed in a two-site distributed call processing model. The overall design concept is intended to mirror a military field unit communicating with a geographically displaced higher headquarters element. Routers, terminals, and software components are consistent with those found in emerging military networks [35]. The testbed occupies three equipment racks (East, Center, West) according to their appropriate position in the deployed network scenarios. All MEU and field unit material resides in the east, data channel simulation at the center, and MEF in the west position. The generic format of the testbed layout is shown in Figure 22.

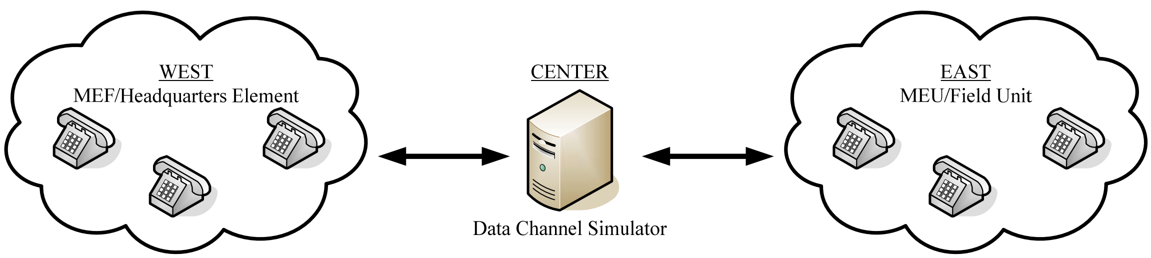

Figure 22. Generic Testbed Layout

The current configuration of the testbed allows for address and hardware expansion to meet future research goals. The remainder of this chapter will discuss the details of existing components and the methods used to connect these individual elements. Figure 23 provides a more detailed view of the testbed topology.

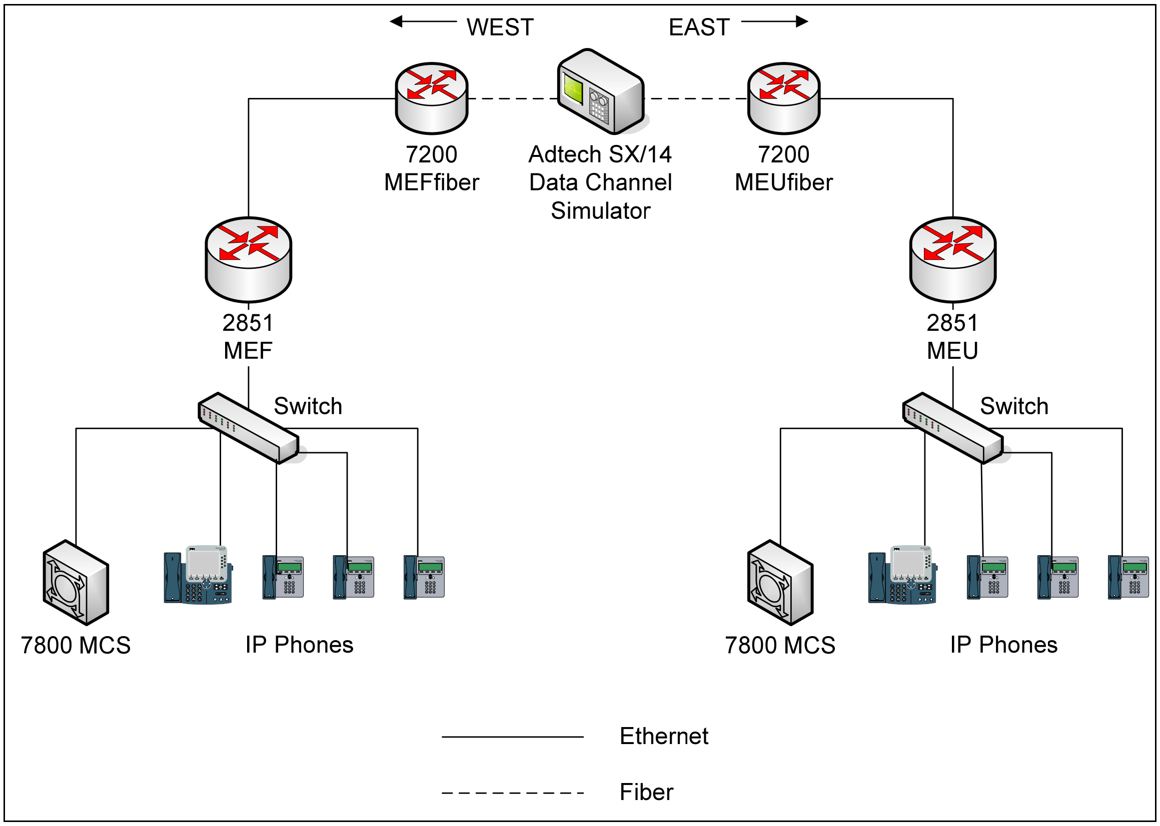

Figure 23. Testbed Hardware Topology

A. COMPONENTS

The elements of the testbed can be traced to the functional components of the H.323 standard. This section will introduce VoIP terminal devices, network control software, and the related physical hardware required to connect and route traffic for experiments.

1. Phones

All VoIP streams require a terminal interface for generation and termination. This testbed uses commercial IP phones, shown in Figure 24, to serve as the end user devices. Operator and maintenance information for each of the Cisco 7911G and 7970G terminals are available in [36, 37].

Figure 24. Cisco 7911G and 7970G IP Phones (from [36, 37])

a. CP-7911G

This terminal represents a mid-level IP phone targeting an office or factory environment. The pixel display promotes user navigation through setting information and call actions. The phone supports eXtensible Markup Language (XML), IEEE 802.3af Power over Ethernet (PoE), G.711 and G.729 audio codecs. All testbed Cisco 7911G phones utilize the PoE option. A built-in data hub allows secondary device access to the parent network. Appendix A explores the device web interface.

b. CP-7970G

This high-end IP phone targets the needs of the business environment. The terminal combines a color touch screen for call function and XML capable web browsing. Additional soft keys are programmable through CallManager and the device settings menu. These phones support PoE, G.711 and G.729 audio codecs. All testbed Cisco 7970G phones utilize the PoE option. A built-in switch allows two secondary device connections access to the parent network. Appendix A explores the device web interface.

2. Cisco 7800 Series Media Convergence Server (MCS)

Each side of the testbed contains a Cisco 7800 series MCS. These units contain Pentium D dual core 2.8-GHz processors, 2 GB RAM, and two removable 80-GB hard drives. These servers store and run all Cisco CallManager 5.0(4) software for the testbed. In addition to their role in regular call processing tasks, CallManager allows these units to be designated as Music On Hold (MOH) servers. This capability allows WAV files to be stored and selectively accessed for playback during voice quality assessment experiments.

3. Cisco CallManager 5.0(4)

Cisco CallManager 5.0(4) acts as the call processing and administrative controller to the testbed device clusters. This software system conducts signaling and call control for the deployed VoIP infrastructure. In large scale VoIP networks a group of servers running CallManager are often joined together to maintain redundancy and call load balancing. In contrast, the testbed design handles a small call load with no bounds on service reliability. Network topology ensures signaling, call control, and voice streams between clusters are subject to the operator defined effects of the test channel. Achieving this objective requires proper understanding of the CallManager administrative features. Four areas of interest to VoIP testing within this network are directory control, codec control, dial patterns, and MOH service.

a. Directory Control

Each terminal device registered to a CallManager receives a directory number allocation through manual or automatic discovery based on the experiment numbering plan. To simplify testing, the network retains only the last four digits associated with the standard North American Numbering scheme. The leading digit is reserved for cluster identification. The three trail digits express the full range of the test clusters. Table 4 shows the CallManager representation of this directory space. X is considered a wildcard digit that can take any value from 0 to 9.

MEF Directory Space	MEU Directory Space
1XXX	2XXX

Table 4. Testbed Directory Range

Table 5 defines the full directory of registered VoIP terminals. During a typical call sequence, structure and range of each cluster's directory drives route pattern matching and codec assignment. Calls established between terminals within the local cluster are said to be on net (e.g., 1000 dials 1001). Conversely, a call that connects to a terminal external to the local cluster is called off net (e.g., 1000 dials 2000).

MEF Device	Directory Number	MEU Device	Directory Number
7970G (CG)	1000	7970G (MEU CO)	2000
7911G (SgtMaj)	1001	7911 (MEU S-1)	2001
7911G (G-2)	1002	7911 (MEU S-2)	2002
7911G (G-3)	1003	7911 (MEU S-3)	2003

Table 5. Testbed Directory Plan

b. Codec Control

Table 6 shows audio codecs and estimated bandwidth consumption for a CallManager handling audio traffic. Standard codec bandwidths are provided for comparison. Actual bandwidth depends on packet size and overhead. The Cisco advertised bandwidth calculations assume 30-ms data packets with IP headers included. A single call is composed of two voice streams. Experiment settings must account for the network capability to carry codecs that are not supported by the VoIP terminal devices. Testbed phone traffic must use G.711u, G.711a, G.729a, or G.729b audio codecs.

Codec	Standard Bandwidth	Cisco Advertised Bandwidth per Call (30 ms packets, IP headers included)
G.711	64 kbps	80 kbps
G.722	48, 56, or 64 kbps	80 kbps
G.723	5.3 or 6.3 kbps	24 kbps
G.728	16 kbps	16 kbps
G.729	8 kbps	24 kbps
Wideband	--	272 kbps
GSM	--	29 kbps

Table 6. CallManager Audio Codecs (after [38])

The CallManager organizes terminal devices associated to a cluster using administrative regions. This approach to call processing accounts for LAN and WAN performance normally associated with geographic separation of network nodes. These parameters are not restricted to true physical location and provide one method for variable codec assignment within the testbed. Figure 25 shows CallManager execution of regional codec controls. Application of this technique is demonstrated in Appendix B.

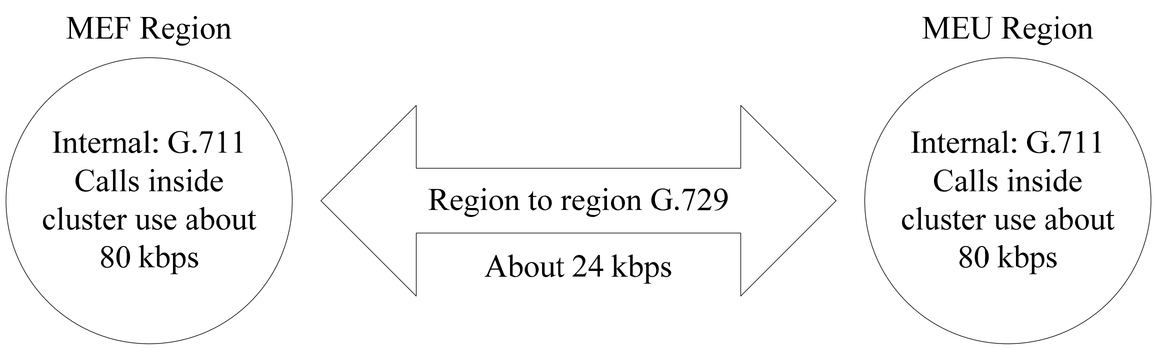

Figure 25. Example of CallManager Regions

c. Dial Pattern Matching

Dial pattern matching helps CallManager recognize a unique group of directory numbers for a specific call processing task. The testbed uses programmed dial patterns to recognize calls that should terminate within, or external to, the local device cluster. These on net and off net calls are processed in a different manner due to the location of registration information. Testbed dial pattern matching and actions are shown in Figure 26.

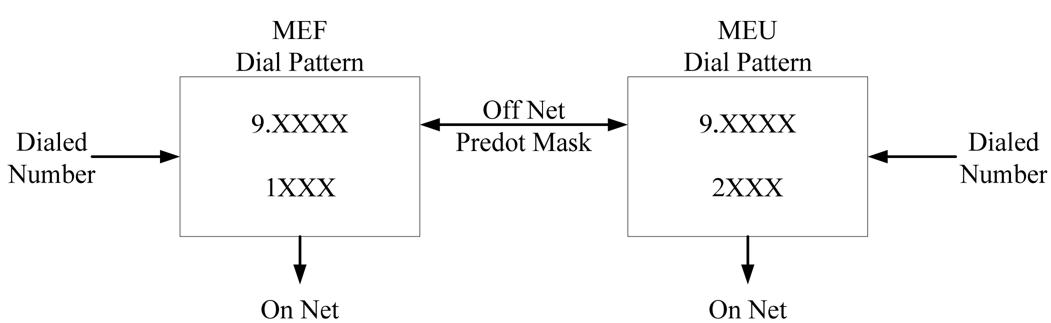

Figure 26. Testbed Number Handling Using Dial Patterns

A call initiated from the MEF cluster to a terminal within the local group of devices (e.g., 1000 dials 1001) only needs call signaling and control services from a single CallManager. A call involving terminals from different clusters (e.g., 1000 dials 2000) requires negotiation between two CallManagers. Testbed dial patterns are associated to a router configured as a H.323 gateway. The dial patterns employ predot functionality for number sequence alteration and handling. Figure 27 shows how the dial patterns function during a sample call.

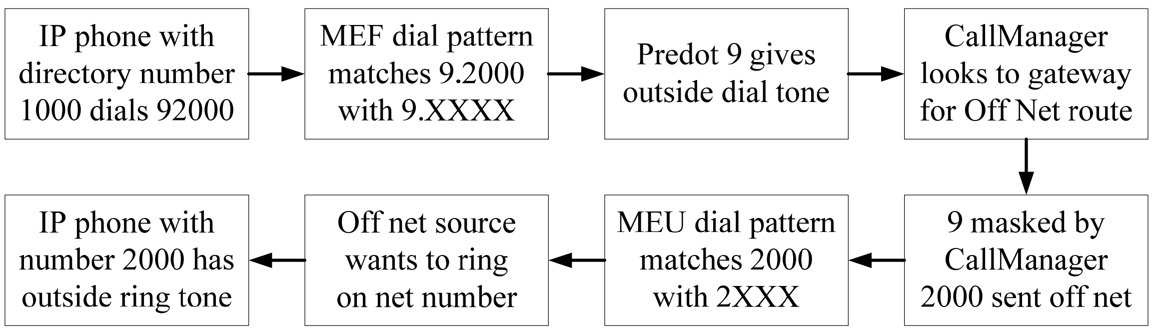

Figure 27. Sample Dial Pattern Actions

d. Music On Hold (MOH)

One noteworthy challenge in telephony testbed design involves repeated uniform injection of a voice input. Variation in background noise from the sender's speaking environment is undesirable when conducting experiments to measure the impact of network channel noise. The testbed overcomes this obstacle by exploiting CallManager's MOH feature. Reference [38] outlines acceptable file formats (e.g., WAV) for this purpose. Sample voice inputs used for this thesis are available from [6] and [39]. These files incorporate the ITU recommended mixture of tempo, active, and passive elements of regular speech. All thesis voice samples contain native English speakers from North America and Europe. CallManager assigns a number and file name to each MOH audio sample. The testbed stores and retrieves MOH for playback by designating the MEF Cisco 7800 series MCS a MOH server. Table 7 displays the codecs supported by MOH playback compared to typical VoIP services. CallManager refers to terminal device or call cluster configuration parameters prior to conducting the signaling for a hold session. The party that initiates a hold session determines the file for playback. Testbed phones point to the desired audio source number for each experiment. A detailed list of instructions for uploading and managing MOH files can be found in Appendix B.

Audio Codec	CallManager	7911G	7970G	MOH Service
G.711	**Yes**	**Yes**	**Yes**	**Yes**
G.722	Yes	No	No	No
G.723	Yes	No	No	No
G.728	Yes	No	No	No
G.729	**Yes**	**Yes**	**Yes**	**Yes**
Wideband	Yes	No	No	Yes
GSM	Yes	No	No	No

Table 7. Testbed Audio Codec Compatibility

Signaling and RTP stream adjustments during a hold session combine to isolate a desired voice exchange for observation. The packet capture graph in Figure 28 reveals a new set of TCS messages in conjunction with a hold session initiation. CallManager closes the logical channel of the first conversation containing undesirable noise. The RTP stream that emerges from a hold session plays a file from the MOH server subject only to desired testbed network effects.

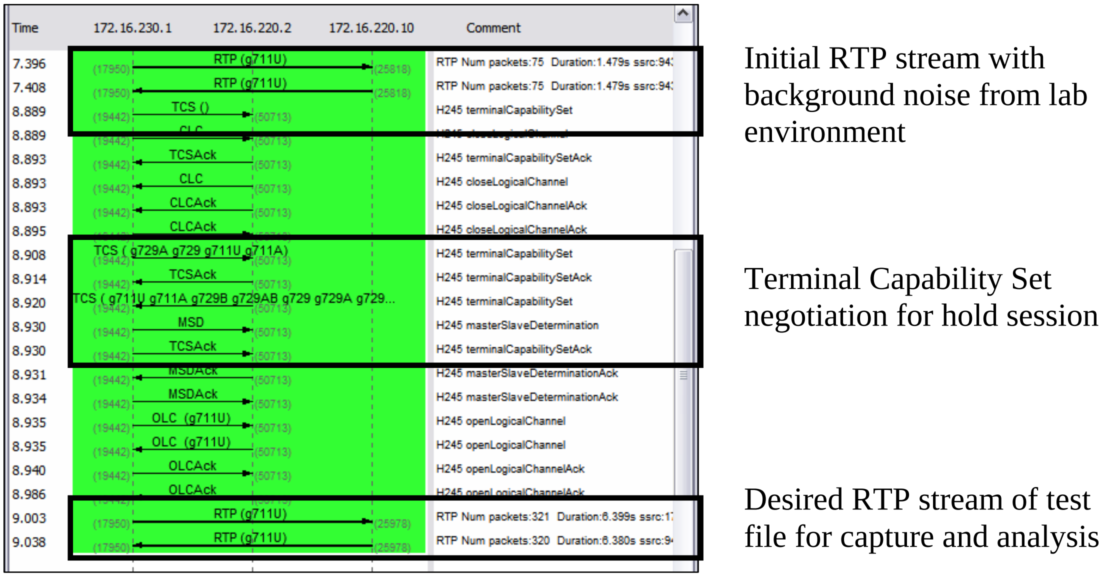

Figure 28. Message Flow During Hold Initiation

4. Netgear FS752TPS Switch

Local call clusters connect to subnet devices using a Netgear FS752TPS switch. Each unit includes 48 10/100 Ethernet ports and 4 Gigabit Ethernet ports. The first 24 ports provide standards based IEEE 802.3af PoE to all testbed IP phones. All port management functions are controlled via a software and web interface. The most current release of switch management software and documentation can be downloaded from the site shown in [40]. The switch provides network connectivity for the phones, MCS, and Cisco 2851 router within each CallManager cluster. Stack management tools enable the switch administrator to monitor all testbed traffic flowing through the device via port mirroring. In this mode, one port is programmed to broadcast transmit and/or receive traffic from any combination of the remaining ports. Port 12 of each chassis was configured to duplicate all switch traffic. These mirror connections facilitate network and call analysis using the open source packet sniffers discussed later in this chapter. Figure 29 is an example of the switch management web interface.

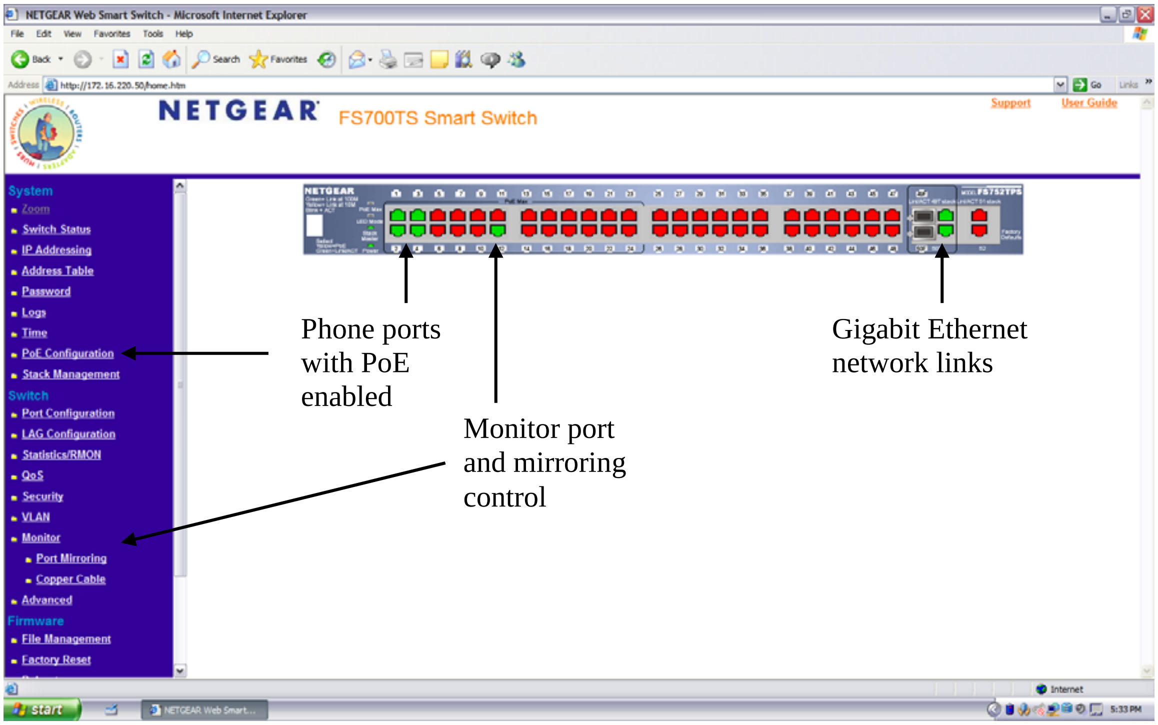

Figure 29. FS752TPS Switch Management Interface

5. Cisco 2851 Router

Call signaling, control, and voice traffic departing a cluster subnet will first encounter a 2851 router. Each 2851 contains two Gigabit Ethernet ports and an IEEE 802.11g capable radio interface. Expansion slots are available to incorporate FXS analog phone input cards servicing two POTS phone lines per Cisco 2851 chassis. Activating the VoIP specific features of each Cisco 2851 required some unique command line inputs. Additional gateway instructions were necessary during the programming of the MEF router. This section addresses the relevant VoIP items encountered during testbed design and construction.

a. H.323 Gateway Configuration

Any attempt to complete inter-cluster calls requires the coordination of both testbed CallManagers. The MEF 2851 router handles the gateway task of negotiating cross cluster H.323 communications. A previous section regarding dial pattern matching linked off net call routing to the testbed gateway. The following lines of the configuration file bind this routing event to a specific port on the gateway.

```
Interface GigabitEthernet 0/1
…
 h323-gateway voip interface
 h323-gateway voip bind srcaddr 172.16.230.1
```

For the case of off net calls departing the MEU cluster, 172.16.230.1 represents the destination port for resolution of call processing tasks involving an external directory number. The gateway receives these requests and forwards H.323 traffic according instructions provided by a dial peer.

b. Dial Peers

Dial peers are similar to dial patterns found in the CallManager setup. Just as the local cluster matches internal or external calls to a pattern, a gateway matches a dialed number sequence to a target IP address. The following configuration lines show a pattern match for calls from the MEU cluster to the MEF cluster. Periods indicate wildcard digits within the dial peer number sequence.

```
Dial-peer voice 10 voip
 description Calls from MEU to MEF
 destination-pattern 2…
 session target ipv4: 172.16.220.2
 codec transparent
```

The session target supplies the CallManager IP address required for further call signaling. Testbed dial peers allow codec negotiation between endpoints. H.245 messages arriving along the dial peer path were formatted using commands within the voice service menu.

c. H.245 Configuration

VoIP service parameters are maintained inside the router H.323 settings. The following configuration file section details voice service elements necessary for testbed voice and MOH operations. Empty capability TCS values must cross the gateway boundary to prevent call disconnect during hold session initiation. Likewise, nonstandard messaging extends service functionality to material covered in [41].

```
voice service voip
 allow-connections h323 to h323
 h323
  emptycapability
  no call service stop
  h245 passthru    tcsnonstd-passthru
```

6. Cisco 7200 Router

The Cisco 7200 series routers that connect the network backbone perform interface and protocol translation required to incorporate the data channel simulator. Each Cisco 7200 chassis contains Fast Ethernet and OC-3 Packet over SONET (PoS) ports. Channel parameters are controlled along the PoS link between each Cisco 7200 router. Testbed data flow and protocol structure are shown in Figure 30. This design enables each router within the testbed to conduct IP routing using OSPF.

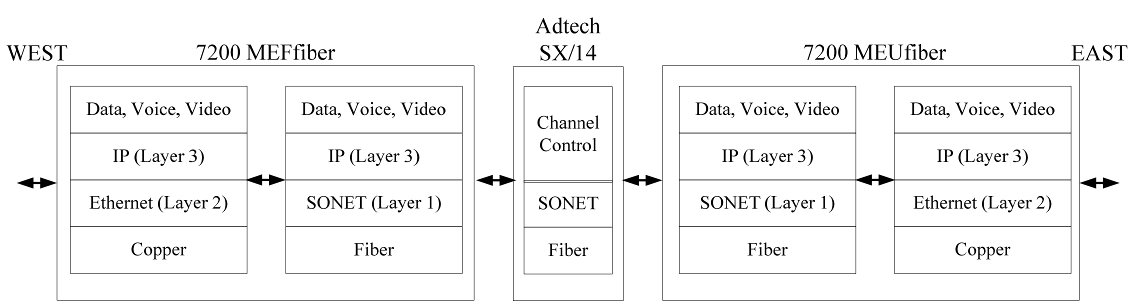

Figure 30. Cisco 7200 Router Interfaces

7. Adtech SX/14 Data Channel Simulator

Configuration of the Adtech SX/14 provides direct control of the testbed channel characteristics. An in depth review of the device is available from [42]. The data channel simulator has been placed in line between two Cisco 7200 series routers. All interfaces operate on a SONET OC-3 155.52-Mbps link. The Adtech SX/14 recovers a clock signal from the MEFfiber router for proper network synchronization. Operator adjustments can be made to delay and error characteristics of the channel. Figure 31 shows a typical data path for traffic inside the simulator. East and West bound traffic represent packets destined for the MEUfiber and MEFfiber routers, respectively. The channel characteristics fall into two categories, delay and error. East and West directed traffic can be controlled independently for asymmetric channel modeling. Custom programs permit multiple combinations of delay and error to run in series. The programming option can string individual channel settings together for a single run or loop the entire group for continuous operation.

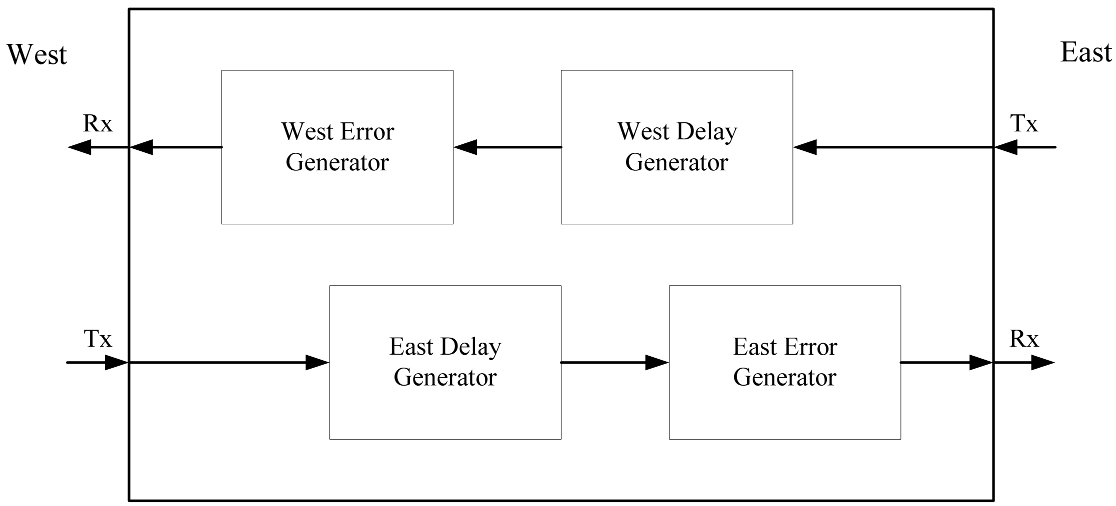

Figure 31. Channel Simulator Data Path (After [42])

a. Delay Control

The Adtech SX/14 uses variable length first-in-first-out delay buffers on each channel. Alterations in the delay program result in recalculation of the delay buffer length. OC-3 connections have a valid delay range from 0 to 324 ms with 1-µs resolution. At data rates of 155.52 Mbps, the buffer can also be selected to a corresponding bit length with 48-bit resolution.

b.   Error Control

Each Adtech SX/14 channel has two error generators that insert logical inversions of transmission data. The first generator is dedicated to the creation of random errors. The second generator provides burst errors. All error distributions are Gaussian. Random error rates can range from 1×10^{-12} to 1 error/bit. Random error injection occurs continuously when no bursts are programmed. In the presence of a burst event, the Adtech SX/14 applies the random error to burst gaps only. Burst programs are set according to error length, error density, and gap length. Valid burst length ranges from 1 bit period to 99,999,999 ms. Burst density determines the error rate within the burst length. Density can range from 1×10^{-8} to 1 error/bit. Gap length determines the time separation from the end of one burst event bit to the start of the following event bit. In the presence of a burst program, the random errors will only be injected during burst gaps. Figure 32 shows a sample of random and burst error generation on the same channel.

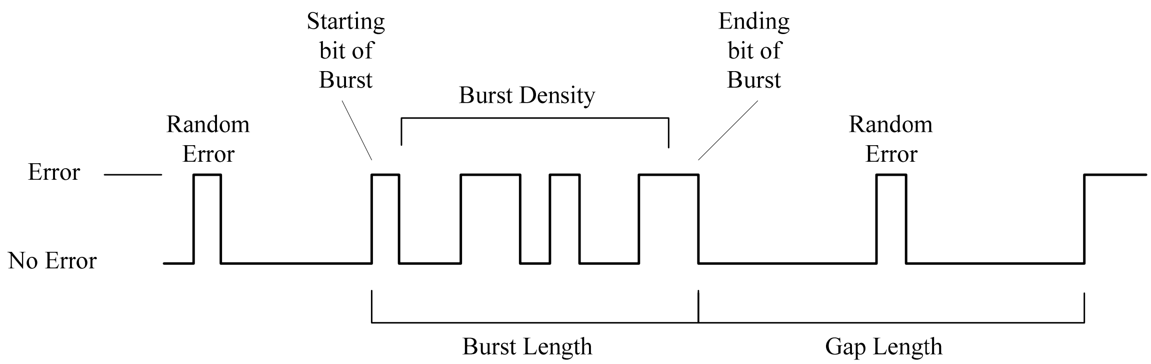

Figure 32. Adtech SX/14 Generated Error Stream (after [42])

B. INTERNET PROTOCOL ADDRESS ASSIGNMENT

All routers within the testbed are configured to network across a single OSPF area. Subnet boundaries are used in a two-layer design architecture. The core area consists of the Adtech SX/14 Data Channel Simulator, Cisco 7200 series routers, and terminates along the Cisco 2851 routers. The access area contains two isolated

CallManager clusters and their associated terminal devices. Figure 33 depicts the general structure of an IPv4 address according to network, subnet, and host identification sections.

N bits	M bits	32 – N – M bits
Network ID	Subnet ID	Host ID

Figure 33. IPv4 Address Structure

Table 8 shows a breakdown of the available address space within current testbed subnets. This scheme provides a simple network hierarchy for data analysis. Address space contained within current subnets is sufficient for potential network expansion.

Location	IP Address Space	Subnet Mask	Subnets Assigned	Assigned Host IDs per Subnet	Remaining Host IDs per Subnet
Core	172.16.230.X	255.255.255.248	3	2	5
MEF	172.16.210.X	255.255.255.0	1	7	247
MEU	172.16.220.X	255.255.255.0	1	7	247
Note: First and last address in subnet range are reserved for net ID and broadcast address respectively					

Table 8. Division of the 172.16.X.X Address Space

Figure 34 illustrates the testbed IP address assignment reflected in routing tables. IP addresses 172.16.210.50 and 172.16.220.50 are designated for the switches associated with the subnet call cluster. A web based device utility allows network administrators to browse and monitor operating status, or configure switch settings. No regular network traffic originates or terminates at the IP addresses.

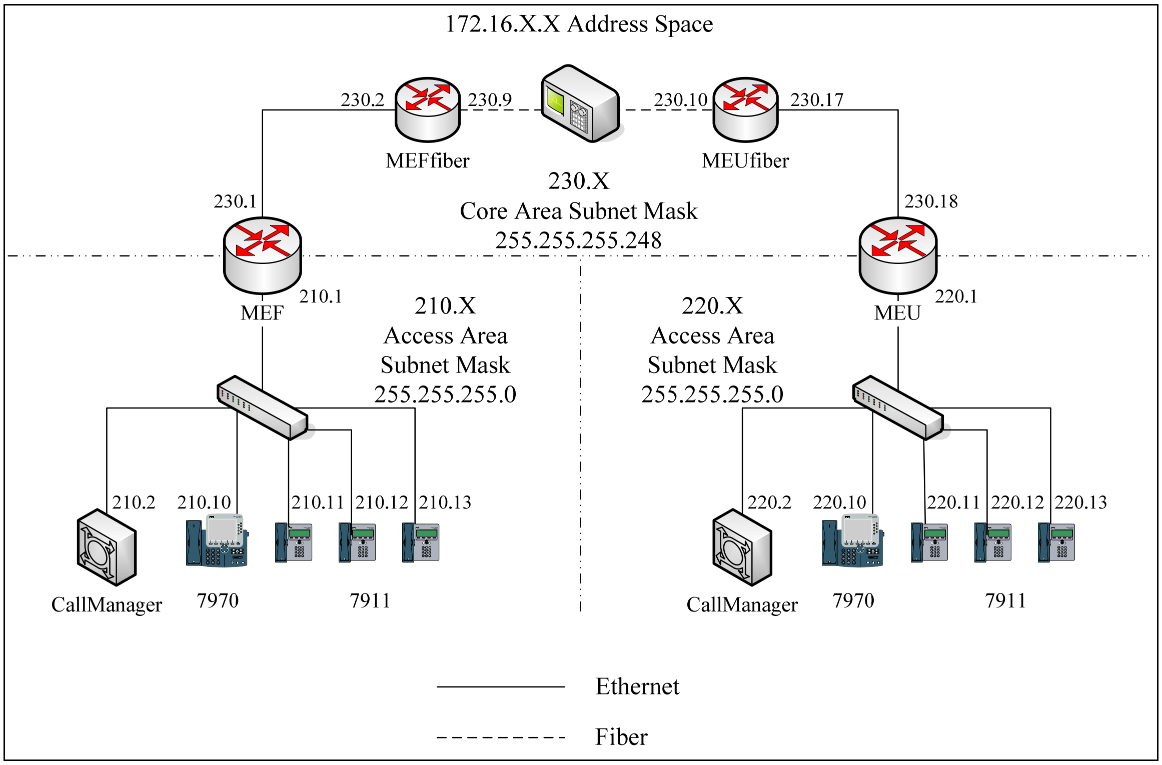

Figure 34. Testbed IP Address Assignment

C. DATA COLLECTION TOOLS

This thesis utilizes a mixture of open source and commercial software platforms for data collection and analysis. The open source material offers a free, flexible alternative to competing network monitor tools. Commercial voice recognition software use is intended to extend and verify previous thesis research conducted at the Naval Postgraduate School. Additional capability within existing network CallManager software was explored for statistical modeling and objective assessment of listening voice quality.

1. Wireshark 0.99.5

Wireshark, formerly released as Ethereal, is the result of an international open source project started in 1998. Program download and reference documentation are available from [43]. The software transforms a normal network interface card into a general purpose traffic monitor. Capture files can then be filtered according to the filters supplied with the Wireshark download. Figure 35 shows a normal testbed traffic capture. The top half of the screen shot provides a list of packet intercepts arranged by time of receipt. The bottom half of the window expands one packet containing H.225.0 call setup information. Hexadecimal content from the H.225.0 packet appears highlighted at the bottom left of the image. This general overview of traffic on the testbed was helpful in detection of initial system configuration errors. Captures at this level still include router management packets interlaced with the VoIP calls. The remainder of this section will focus on Wireshark VoIP statistic options used to extract speech information from packet capture files.

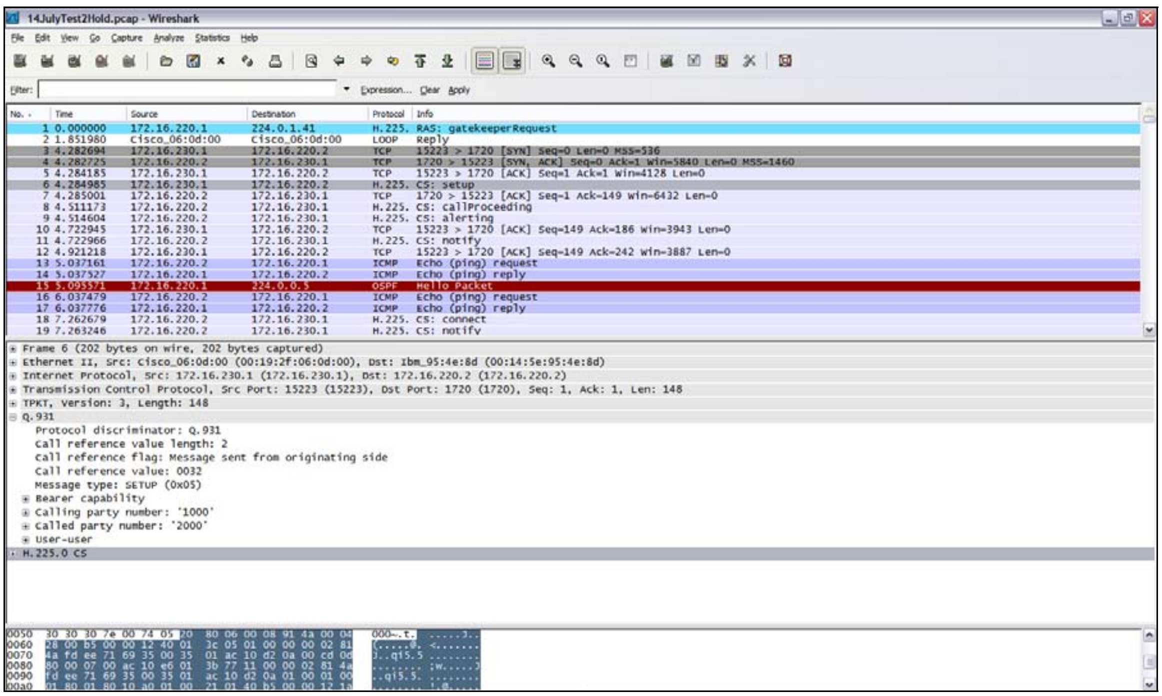

Figure 35. Wireshark Packet Capture with Expanded H.225.0 Message

Wireshark includes a tool for the filtering and deconstruction of any captured H.323 or SIP exchange. Signaling messages are linked to the subsequent RTP streams for graphical display and decoding for playback. Figure 36 shows the timeline analysis of an H.323 call. The player has already decoded the voice traffic for playback using the variable jitter buffer setting of 20 ms. Valid Wireshark jitter buffer range includes values from 0 to 50 ms in 1-ms increments.

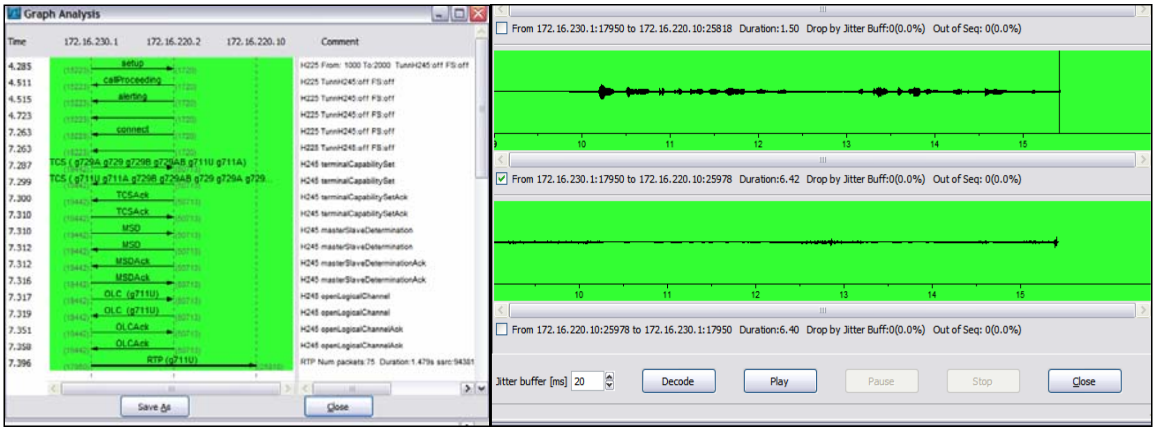

Figure 36. Wireshark VoIP Call Graph Analysis and RTP Player

The VoIP statistic options are limited to calls between testbed CallManager clusters. Internal cluster calls do not require an H.245/225.0 exchange since a single call manager conducts all processing. In these cases, Wireshark does not detect an H.323 event for decoding as a VoIP call. External calls are intercepted as an H.323 event, but decoded voice playback requires Wireshark's RTP player. The constraint on voice file export format led to the testbed assimilation of another open source software tool.

2. Cain and Abel v4.9.1

The Cain and Abel pair of programs originally emerged as a password recovery utility for computers running Microsoft operating systems. Updated versions have expanded the capability for the Cain half of the software package to probe network routing protocols and record VoIP conversations in a WAV format. Testbed call intercepts use Cain in a two step process. Upon initial connection to the network, via a Netgear switch, Cain conducts topology mapping and an ARP Poison Routing (APR) routine. This step manipulates host ARP caches to conduct a form of man in the middle hack. Figure 37 illustrates regular and APR enabled routing of VoIP packets between a MEU and non-MEU phone.

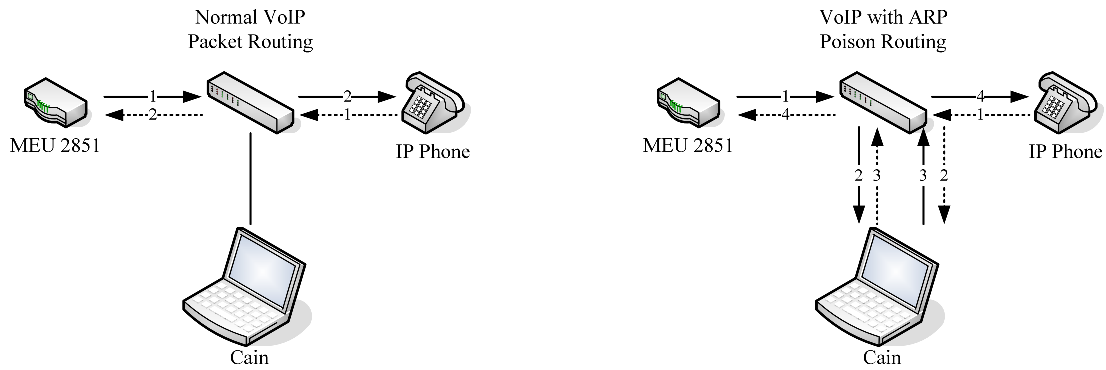

Figure 37. Cain ARP Poison Routing

Following the manipulation of router and host phone ARP cache, Cain silently intercepts the VoIP RTP stream for recording. The second step isolates the desired RTP from the VoIP session for decoding and WAV file construction. A single VoIP call within the testbed may result in multiple RTP streams based on the use of hold sessions or conference call options. WAV files generated for analysis in this thesis are restricted to mono output format for speech to text conversion. Figure 38 shows the appropriate Cain recording window. Product download, supported codecs, and detailed instructions for using the technique described in this section are available from [44].

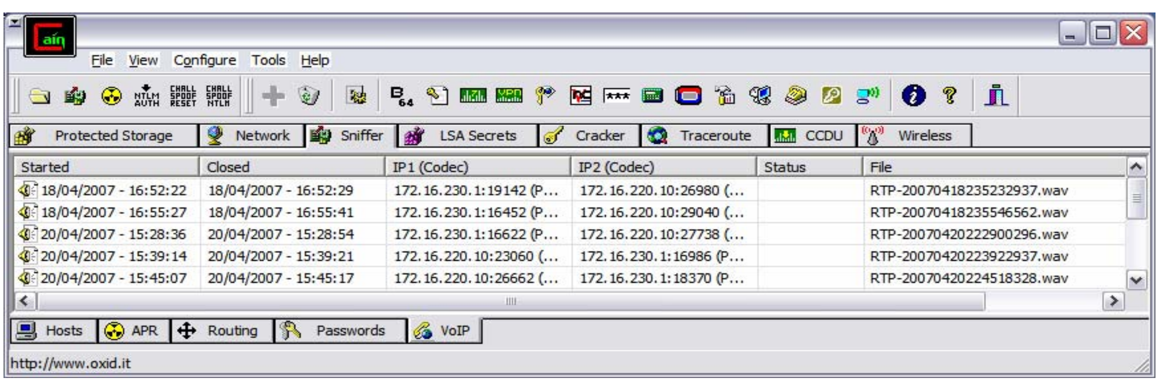

Figure 38. Cain VoIP Recorder

3. Dragon NaturallySpeaking 9.0

Dragon NaturallySpeaking is a voice recognition software product produced by Nuance Communications. Available background material on the specific techniques exploited by Nuance engineers is limited to [45]. Tiantioukas examined the suitability of using commercial voice recognition software in the estimation of VoIP voice quality [9]. This thesis extends the approach established by Tiantioukas to the testbed environment described in this chapter.

The accuracy of voice recognition software improves with the initial training and subsequent use. Corrections to translation errors also assist the software in improving translation quality. A review of the product documentation suggests a Hidden Markov Model approach to voice recognition is used by NaturallySpeaking. Testbed software initial training was conducted per device installation instructions for a new user. WAV files recorded from Cain packet captures were processed through the Dragon speech to text translator. No attempt was made to improve long term accuracy through text translation error correction. Control files were generated by setting all data channel injected error levels to zero.

4. Cisco Call Statistics

Cisco IP phones have the ability to display a series of voice quality statistics compiled during the course of an established RTP stream. Appendix A describes each element within the statistics table obtained from a Cisco 7970G web interface. Cisco phone documentation [46] defines three key parameters: concealment ratio, concealed seconds, and MOS-LQK. When an RTP stream sent to an IP phone suffers frame loss, a concealment frame is inserted by the digital signal processor (DSP) to mask the event. The concealment ratio is given by

$$\text{Concealment Ratio} = \frac{\text{Number of concealed frames}}{\text{Total number of speech frames}} \tag{4.1}$$

where the concealed frames are calculated in three-second intervals. Any one-second interval containing a mask frame from the DSP increments the concealed seconds counter. Single second intervals including more than five percent masking are considered severely concealed. A proprietary algorithm developed by Cisco computes these metrics in a continuous fashion for the previous eight second window to calculate the MOS-LQK. This objective assessment of voice quality is consistent with ITU provisional standard P.VTQ.

D. SUMMARY

In this chapter, a testbed design for non-intrusive objective voice quality assessment was introduced. Detailed control of the network data channel includes error and delay metrics. Finally, data capture and analysis tools were presented for extended application to thesis testbed experiments.

THIS PAGE INTENTIONALLY LEFT BLANK

V. TESTBED EXPERIMENTS

The experimental results presented in this chapter were generated through the evaluation of approximately ten hours worth of voice file transmission across the testbed VoIP network. Individual test runs were carried out using one minute data collection periods. Call statistics for each run were transferred to Matlab for collective analysis and plotting. Voice files were captured and transferred to voice recognition software for subsequent clarity analysis. Figure 39 shows the typical sequence of events required for each data run.

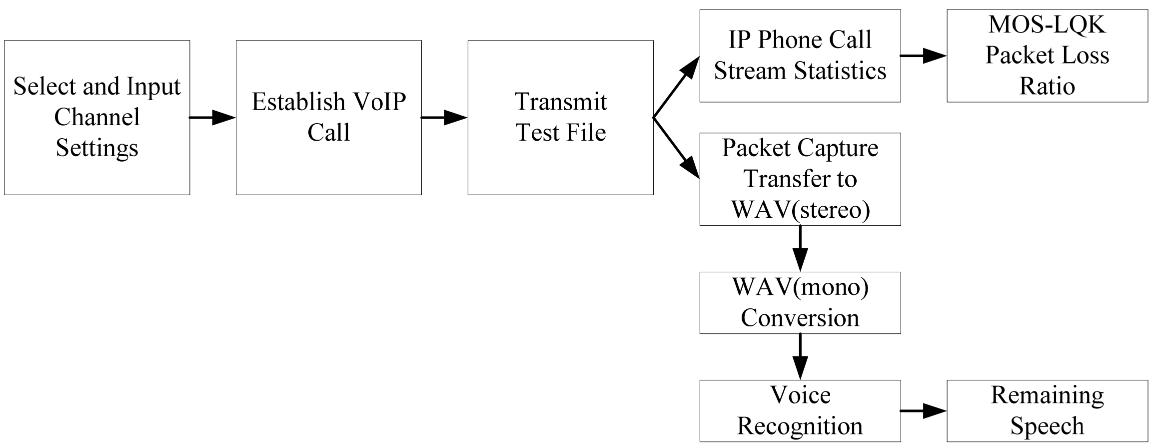

Figure 39. Experiment File Transmission and Data Collection Sequence

Network statistics of interest included the bit error rate (BER), packet loss ratio, and MOS-LQK. BER, commonly used as a metric in the performance evaluation of communication systems, is given by:

$$\text{BER} = \frac{\text{Number of bits received with error}}{\text{Total number of transmitted bits}} \tag{5.1}$$

Occasionally, network effects resulted in the failed delivery of entire packets. A useful mathematical representation for evaluating these events is the packet loss ratio, given by:

$$\text{Packet Loss Ratio} = \frac{\text{Packets transmitted} - \text{Packets received}}{\text{Packets transmitted}} \tag{5.2}$$

The remaining metric, MOS-LQK, is recovered directly from the Cisco 7970G phone terminal at the conclusion of each test. To quantify the impact of BER and packet loss on received speech comprehension, this thesis uses the concept of remaining speech from [9]. Using voice recognition software, calls captured at the receiver side of the testbed via Cain are transcribed from WAV file format into a text document. Text conversion is reviewed for translation accuracy. Runs are then compared to the output text with the channel simulator error injection set to zero. Remaining speech is calculated by

$$\text{Remaining Speech} = \frac{\text{Number test file words transcribed correctly}}{\text{Number of control file words transcribed correctly}} \tag{5.3}$$

A. TESTBED LIMITATIONS

The first series of experiments established valid operating boundaries for remaining data collection runs. Different combinations of BER, delay, and test files were used in an effort to stress the network to failure. Limitations were documented in the area of BER, delay control programs, and voice recognition capability.

1. BER

Random error injection from the channel simulator serves as the principal factor for replicating conditions found in tactical wireless links. The PoS interface used to mimic radio connections is limited by the BER monitor used to evaluate link status. This results in a reduction of the acceptable BER dynamic range available for testing. Observation of the link status alarms along the PoS connection confirmed SONET loss of signal (SLOS) and SONET loss of frame (SLOF) thresholds at a BER of 3×10^{-5}. Crossing the SLOS or SLOF threshold triggered a link status alarm that causes each Cisco 7200 router to disable the PoS link. These actions are intended to evaluate the link for proper physical connection and the suitability of the fiber optic cable. During a failed PoS link period, test calls in progress lost all active RTP streams. No call signaling messages are exchanged with terminals at the point of link failure. Open logical channels void of traffic are observed as each IP phone sat idle with no voice output. Call progress clocks on terminal displays continued to count up. A subsequent reduction in channel simulator BER recovered the RTP connection between phones. Call statistics at each terminal show no packet transfer and a default MOS-LQK of 2.0 during the failure window. Burst error test runs with burst density equivalent to the previous random error parameters revealed matching limitations. The restriction in RTP transfer eliminated the channel simulator BER range of 3×10^{-5} to 1×10^{-3} from further experiments.

2. Delay Programs

The simulation of channel delay characteristics includes both path delay and jitter. Ping test packets traversing the network indicate channel simulator settings are consistent and accurate to ± 1 ms in the reproduction of end-to-end delay. The ability to produce and control jitter within the channel was explored through the use of channel delay programs. Adtech SX/14 channel program features cycled through a series of channel conditions in loop format. The delay profile was set to dwell on different values at irregular intervals in an attempt to create jitter within the network. Observation of the PoS link revealed SLOS and SLOF alarm indications triggered by each program step. Each alarm event propagated a link failure between the Cisco 7200 routers. These alarm events were associated to the time required for the channel simulator to recalculate the new buffer length for the corresponding delay program step. During the calculation interval, a series of logical spaces or marks must be transmitted by the channel simulator. Both of these choices resulted in temporary PoS link failure. These observations limited the use of channel simulator delay to a single setting. In this mode, there is no associated control of jitter within the testbed.

3. Voice Recognition

The voice recognition software used in this thesis requires an interactive training process with a user. Operator profiles are saved within the Dragon NaturallySpeaking software for reference during all dictation or transcription processing events. This thesis used two voice recognition profiles from North American native English speakers (male and female). All software user options for training the profile were disabled or bypassed following initial configuration. Of the four voice files used for testing in this thesis, two contain voice samples of European native English speakers. Transcription attempts for captures from these European speakers failed to provide sufficient material needed to extract associated values for remaining speech. Remaining speech results reported within this thesis are the product of multiple captures of the North American speaker files subjected to various channel conditions.

B. OBJECTIVE VOICE QUALITY TESTS

This section presents the results of testbed experiments obtained from the transmission of speech files using the restricted range of suitable channel settings. BER settings for detailed examination were selected from an evaluation of MOS-LQK and packet loss observed during initial network stress tests. Additionally, these channel conditions were intended to provide a range of data points where degraded testbed voice reception could be analyzed. A summary of test parameters follows:

- Test files: European Female, European Male,
  N. American Female, N. American Male
- Codecs: G.729, G.711u
- Channel BER: Random error ($1\times10^{-6}, 5\times10^{-6}, 8\times10^{-6}, 1\times10^{-5}, 2\times10^{-5}$)
  Burst errors disabled
- Channel delay: 0 ms, Programs disabled

1. MOS-LQK Results

The first data runs examined the effect of channel BER on MOS-LQK values obtained from IP phones receiving a test file. The results from G.729 transmissions are depicted in Figure 40. All test files displayed strong correlation throughout testing. To improve readability of plots, only results for the European Female and North American Male files are provided for remaining graphics in this chapter. Additional test results for G.711 transmissions are shown in Figure 41. A composite view of MOS-LQK results for both codecs for N. American male and European female is shown in Figure 42. The results are based on 15 Monte Carlo runs.

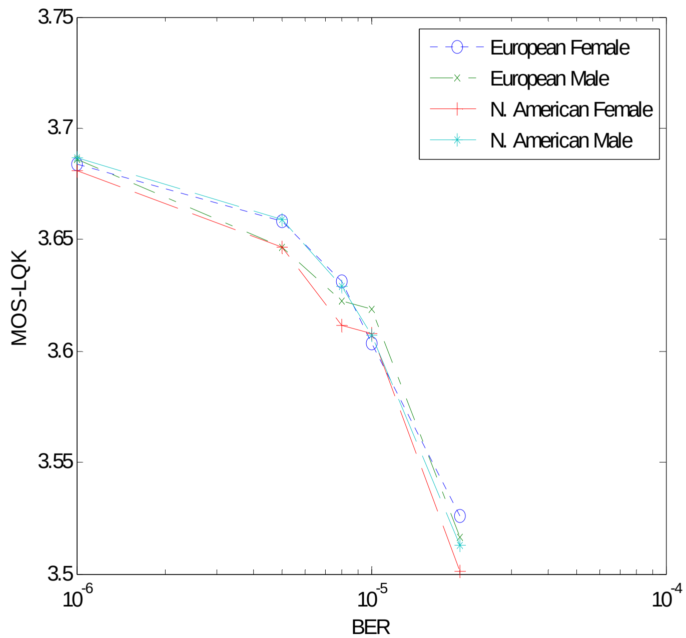

Figure 40. MOS-LQK as a Function of BER for G.729 based on 15 Monte Carlo Runs

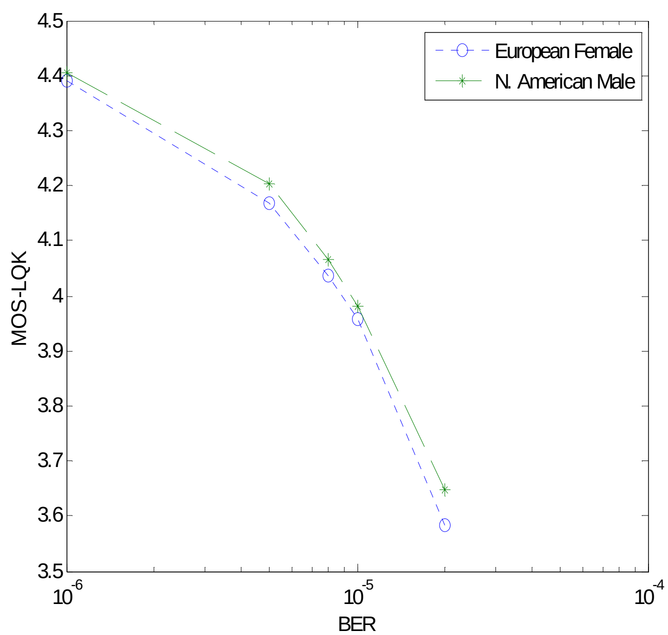

Figure 41. MOS-LQK as a Function of BER for G.711 based on 15 Monte Carlo Runs

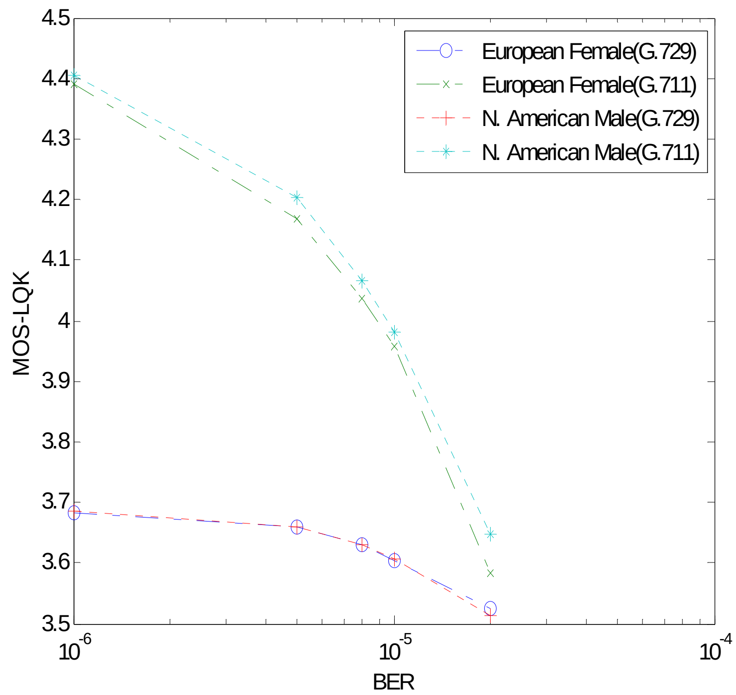

Figure 42. MOS-LQK as a Function of BER for G.729 and G.711 for N. American Male and European Female based on 15 Monte Carlo Runs

As the channel BER rate increases, each codec suffered a corresponding decline in MOS-LQK value. Peak MOS-LQK value for G.729 codec traffic was limited to 3.7 by the Cisco listening quality algorithm. A similar restriction is placed on G.711 MOS-LQK with values capped at 4.5. The testbed capability to degrade G.729 listening quality scores was limited to less than a 0.2 deflection from maximum performance. The corresponding decay in G.711 testing registered an approximate 0.95 reduction from the maximum score. G.711 managed to provide superior MOS-LQK performance for all data points other than the most severe BER available to the testbed. Similar MOS-LQK trends were observed across all four test files.

The decline in MOS-LQK corresponding to the increased BER is examined further. H.323's use of RTP results in the delivery of individual bit errors contained within the payload of voice packets. The successful transmission of corrupted voice samples has a detrimental impact on the perceived content of speech beyond the scope of

MOS-LQK. MOS-LQK values only focus on the ability for the DSP to transmit frames related to delivered packets. A more destructive event to MOS-LQK occurs when the channel bit error strikes VoIP packet headers. Errors of this nature lead to packet loss, and an increase in DSP concealment frame transmission. Thus, plots of MOS-LQK versus BER show a negative trend that should be corroborated by packet loss data. Likewise, successful frame transmissions in the presence of higher BER require further analysis to quantify the perceived value of speech content. The next two sections address these concerns.

2. Packet Loss Results

After measuring the effect of BER on MOS-LQK values, data points were examined for packet loss impact on MOS-LQK. The results of that analysis are illustrated in Figures 43 and 44 for G.729 and G.711, respectively. Figure 45 provides a composite view of codec data. All plots are based on 15 Monte Carlo runs.

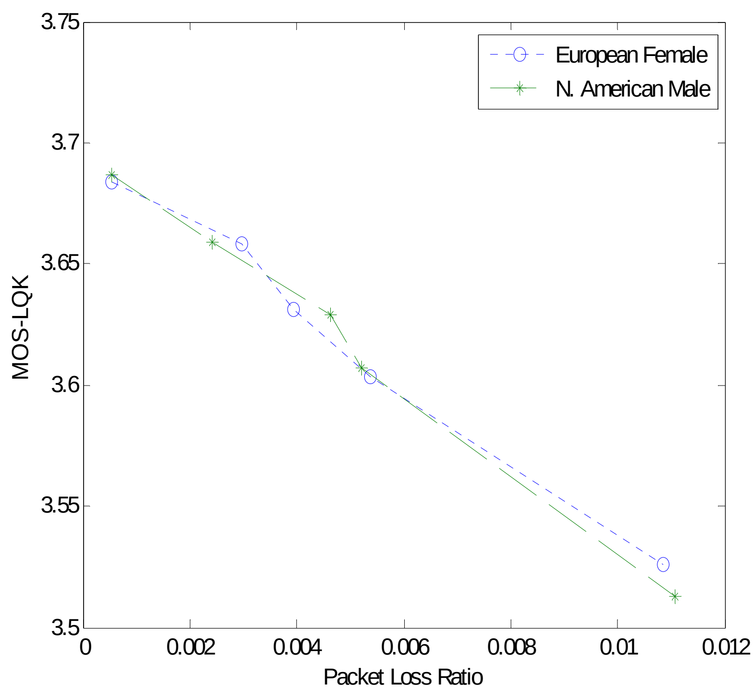

Figure 43. MOS-LQK Ratio as a Function of Packet Loss for G.729 based on 15 Monte Carlo Runs

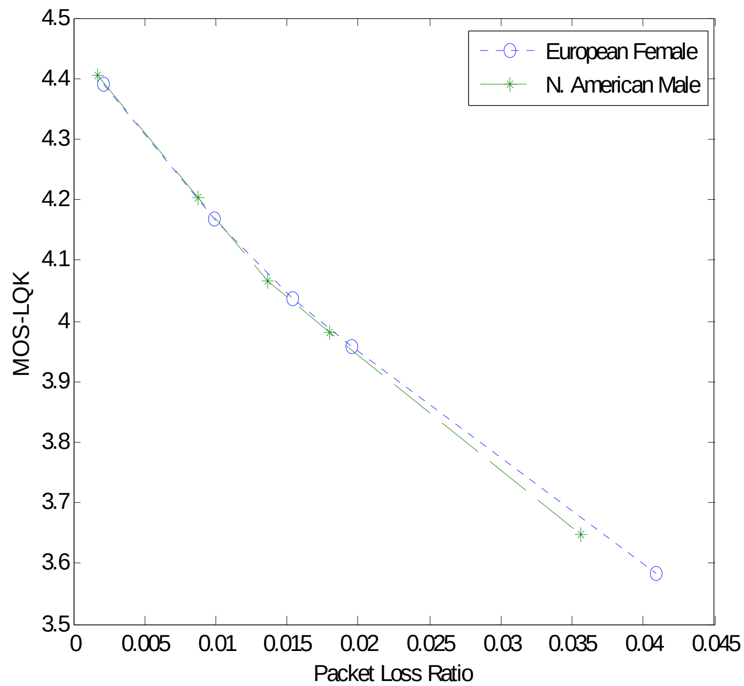

Figure 44. MOS-LQK as a Function of Packet Loss Ratio for G.711 based on 15 Monte Carlo Runs

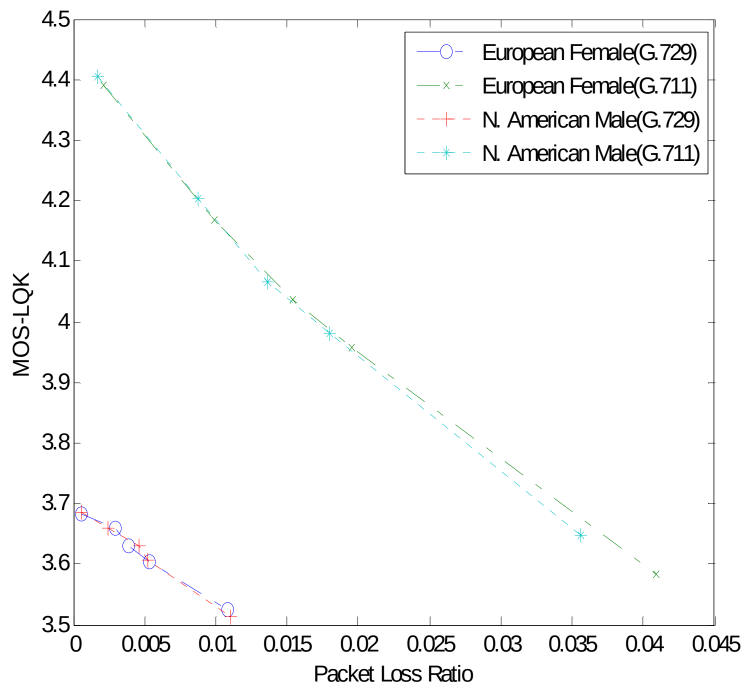

Figure 45. MOS-LQK as a Function of Packet Loss Ratio for G.729 and G.711 for N. American Male and European Female based on 15 Monte Carlo Runs

Analysis reveals a decrease in MOS-LQK consistent with the increase in lost packets for both codecs. G.729 tests suffered less overall packet loss compared to G.711 runs. G.711 MOS-LQK scores outperformed G.729 despite greater packet losses. All test files exhibited similar loss characteristics within each codec family of data points.

The packet loss trend supports BER results with a near linear increase across all test points. MOS-LQK values in this area of packet loss decline in response to the DSP concealment frame compensation for lost voice data. While these tests show a narrow region of packet loss (0 to 4.5 percent), the related rate of MOS deviation is consistent with other objective prediction model calculations [33]. Variations of MOS-LQK value in localized regions of packet loss ratio value can be attributed to the distribution of concealment frame transmissions. Concealment frame bursts resulted in severely concealed segments of an RTP stream with greater impact on MOS-LQK values. Evenly spaced concealment produced less severe deviations in MOS-LQK. The dynamic range of testing was limited by SONET link alarms. Observed losses are specific to channel conditions and do not account for the packet loss VoIP networks experience due to congestion and jitter.

3. Remaining Speech Results

The results in this section explore the impact of BER and packet loss on the amount of comprehensible speech received by the endpoint terminal. Figure 46 presents the amount of remaining speech compared to channel BER. Figure 47 illustrates remaining speech as a function of packet loss. Figure 48 shows plots illustrating the amount of remaining speech as a function of codec and MOS-LQK value. All plots are based on 15 Monte Carlo runs.

BER and packet loss affected the value of remaining speech differently according to the selection of the test file codec. Overall, G.711 outperformed G.729 in analysis of speech intelligibility for the given channel conditions. No loss in content was observed for G.711 until it was subject to the two highest amounts of channel error available. In contrast, G.729 shows immediate reduction in remaining speech. Loss factors associated with G.729 data were amplified due to the compression techniques applied by the codec.

The corruption of bits within packet payloads using G.729 influenced a larger portion of the RTP stream due to errors within a G.711 payload. In general, compressed speech was more susceptible to degradation in intelligibility.

Test file transmissions provide 150 to 180 words for transcription. The average amount of speech lost to the worst case G.729 trial was five percent. This represents three seconds of speech loss per minute, or seven words of the total test file. The G.711's worst case scenario suffered a three percent loss in comprehensible speech. This loss corresponds to roughly two seconds per minute, or four words per test file run.

Disparities were observed between voice recognition of the male and female speakers. These differences can be attributed to the quality of initial software training and individual test file data content. Voice recognition profiles used in this thesis are independent and gender specific. The male voice profile provided a more accurate transcription of the control file. Efficient software training, coupled with higher speech content in test files, helped skew any remaining speech data comparison in favor of the male speaker. Since female test files contained seventeen percent less speech activity, they are more sensitive to word loss given an equal period of observation. Remaining speech observations can be improved through the translation of multiple test files for each independent user. Large scale intelligibility trends related to BER, packet loss, and MOS-LQK are still visible in light of these limitations.

Analysis of remaining speech revealed an important distinction between the perception of VoIP listening quality, measured by MOS-LQK, and intelligibility. Files captured at lower MOS-LQK scores still managed to deliver near perfect remaining speech results. G.729 with a MOS-LQK of 3.7 provided superior comprehension to the listener when compared to G.711.

The experiment identified a tradeoff between bandwidth and performance that often challenges VoIP network design. In regulating the VoIP bandwidth, an administrator directly impacts the quality of speech provided to the receiving party. However, the cost associated with a less accurate reconstruction of human voice does not necessarily deter a listener from extracting useful information during a conversation.

More simply, a person can sound bad while accurately conveying their thoughts. This subtle point is illustrated by the disparity in G.729 and G.711 results. These observations also highlight the importance of establishing a broad concept of performance. MOS-LQK and intelligibility are measures of effectiveness that should be approached as symbiotic elements. Analysis in isolation provides a conflicting and incomplete assessment of the call experience.

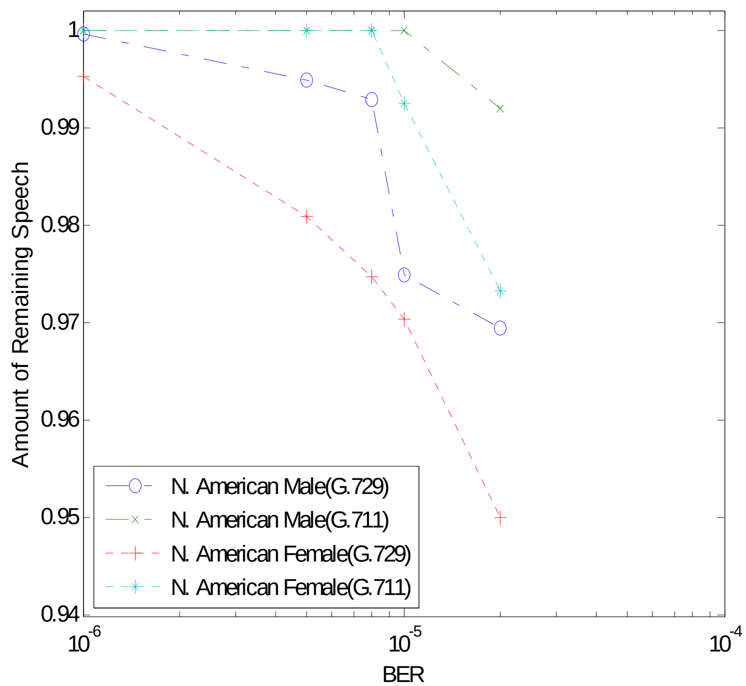

Figure 46. Remaining Speech as a Function of BER for G.729 and G.711 based on 15 Monte Carlo Runs

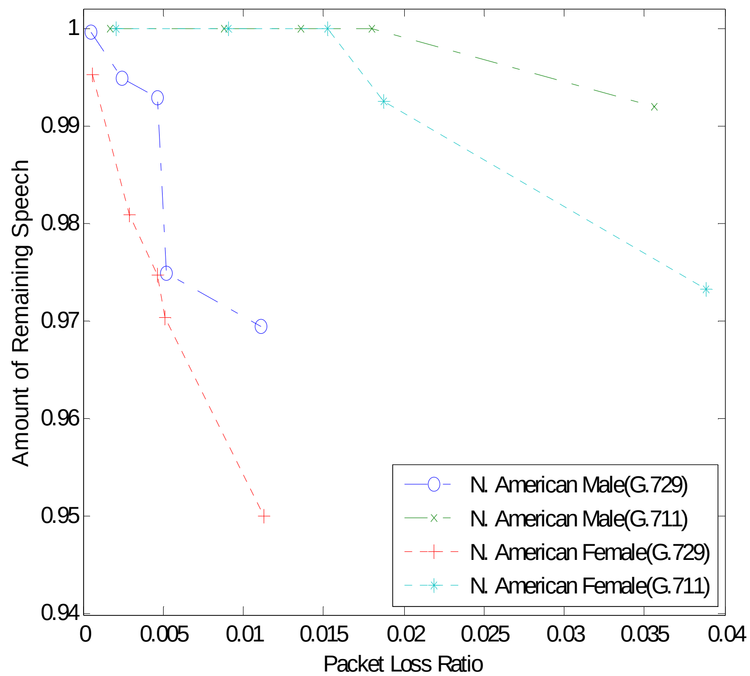

Figure 47. Remaining Speech as a Function of Packet Loss Ratio for G.729 and G.711 based on 15 Monte Carlo Runs

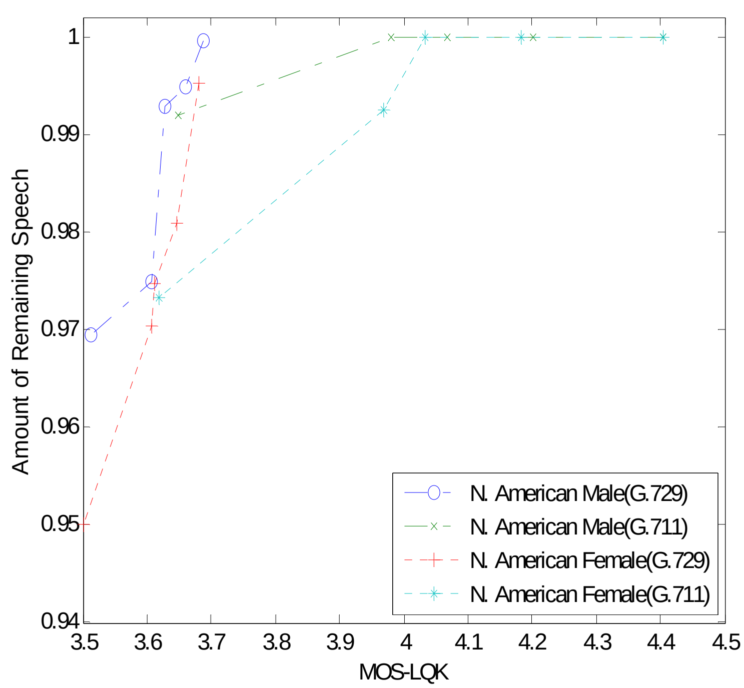

Figure 48. Remaining Speech as a Function of MOS-LQK for G.729 and G.711 based on 15 Monte Carlo Runs

4. Delay Considerations

End-to-end delay provides significant influence to perceived quality of two-way VoIP conversations. MOS-LQK, by definition, only provides mapping of MOS estimates through the analysis of packet loss statistics and DSP activity. Predictive quality modeling, introduced in Chapter III, accounts for the effect of delay when calculating conversational quality estimates, MOS-CQE. This section provides a method for analytically incorporating channel delay forecasts into testbed MOS-LQK data.

The network planning tool, known as the E-model, collects the additive contributions of network characteristics into the R factor defined by Equation (3.1). Experimental MOS-LQK results can be transformed into corresponding R values using Equation (3.5). If we assume that all network conditions other than delay remain unchanged, the R factor can be adjusted by calculating the I_{dd} shift from Equation (3.3). These updated R values blend objective observations with forecast delay considerations. Converting the adjusted testbed results back to expected MOS with Equation (3.5) completes the extension of testbed experimental results to include the effect of delay.

Figure 49 illustrates the application of predictive model adjustments to experimental results. The plot shows estimated MOS for 200, 300, and 500-ms delays in the G.711 North American Male speaker file. The maximum 500-ms delay corresponds to a geosynchronous satellite link round trip. The plot indicates a near linear degradation of experimental results to expected MOS for delays in the range from 150 to 500 ms.

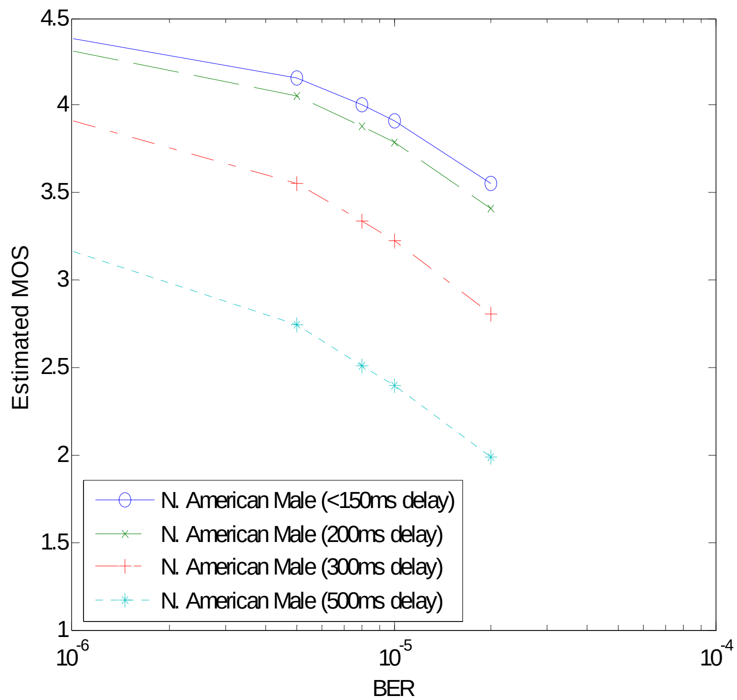

Figure 49. Estimated MOS with E-model Delay Factor Correction as a Function of BER based on 15 Monte Carlo Runs

C. SUMMARY AND DISCUSSION

This chapter presented the results of experiments conducted on the VoIP testbed for the objective assessment of VoIP quality. Limitations of the testbed were identified to establish a valid operating range for the experiments. A sequence of test call results was presented using observations and calculation of metrics to include MOS-LQK, packet loss, and remaining speech. Results were compiled and displayed using MATLAB. Testbed channel simulations demonstrated the controlled degradation of VoIP traffic using either the G.729 or G.711 codec. An approach to incorporate channel delay through predictive modeling was also provided.

Future implementation of tactical VoIP will clearly require more in-depth research and development. Current testbed channel simulations are based upon an imperfect SONET based representation of the wireless environment. Each experiment provides a stepping stone for the evaluation of voice traffic in emerging VoIP networks.

As VoIP penetrates the military market, the typical metrics tied to commercial success may be incongruent with the needs of our deployed forces. Military users are likely to value intelligibility over the fidelity of voice reconstruction. Long delays may be tolerated for service to remote locations. Codec selection, network effects, and conversational comprehension are elements best utilized in a holistic review of VoIP performance. The testbed experiments described in this chapter provide a flexible platform for further exploration of VoIP voice quality characteristics in expeditionary scenarios.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSIONS

This thesis explored the standards used to field VoIP applications. An ITU-T H.323-based VoIP testbed was constructed using Cisco routers, servers, IP phones, Netgear switches, and the Adtech SX/14. Cisco CallManager provided call processing functions through a network monitored by Wireshark and Cain packet capture tools. Dragon NaturallySpeaking supplied voice recognition capability for an examination of speech intelligibility. Additional metrics of BER, packet loss, and MOS-LQK were recovered during test calls using voice files from speakers of both genders and mixed nationality.

Experiments provided results consistent with a conceptual approach to voice quality parameters that defined delay, echo, and clarity. ITU-T subjective, objective, and predictive modeling tools were used to provide voice quality results consistent with telecommunications industry standards. Experiments investigated the testbed’s capability to control VoIP performance through channel simulation and delay prediction.

A. CONTRIBUTIONS

This thesis accomplished two objectives. The VoIP network established for experimentation provides a modern H.323 VoIP research platform. Inherent scalability and flexibility of the design delivers a reusable foundation for future research efforts. The call processing software and the address scheme accommodate potential expansion of terminal device population and diversity. Testbed network design also maintains a topology suitable for rapid reconfiguration. Any alterations at the core area of the design preserve the work previously devoted to call cluster development and programming. Data channel simulator interfaces are isolated and positioned for prospective hardware upgrades.

The testbed successfully facilitated the controlled degradation and measurement of voice quality. Experiments and analysis explored in this thesis provide a cost effective approach to non-intrusive, objective voice quality assessment. These techniques leverage the benefits of open source monitoring tools while extending the use of commercial software for speech intelligibility measurement. Observations indicate that network error management capabilities will be preserved throughout basic design alterations. Delay consideration limitations were overcome through the adaptation of ITU-T E-model delay impairment factor calculations.

B. FUTURE WORK

This study was based on observations of voice quality metrics taken from a H.323 VoIP testbed incorporating the Adtech SX/14 Data Channel Simulator for error and delay control. The current testbed design exhibits some constraints and limitations open for improvement and future research opportunities.

The network described in this thesis used minimal overhead and security settings during the transmission of voice traffic. All components are isolated from outside data exchange and typical patterns of daily human interaction. These conditions result in a level of artificiality that must be acknowledged. True military networks must incorporate security policies while managing the balanced QoS necessary to parse capacity among data and voice needs of the warfighter. While this work has emphasized H.323 connections, future research should consider the incorporation of SIP based services as well.

Some limitations imposed on the testbed are a product of the hardware available for network design. The channel simulator, and associated PoS interface, introduced the primary limitations for experiment parameter range. Current BER dynamic range, delay programming, and jitter control capability establish bounds on the range of channel characteristics for experimentation. A more robust channel simulator and interface would help expand the design beyond PoS link failure restrictions. Future designers altering the testbed should investigate the ability to establish an IEEE 802.11 or 802.16 bridge between the Cisco 2851 routers. These RF links can connect to Spirent 5500 channel emulator according to the proposed network layout in Figure 50. Such an enhancement would allow VoIP testing over a long distance wireless link while providing in-depth control over the channel fading environment.

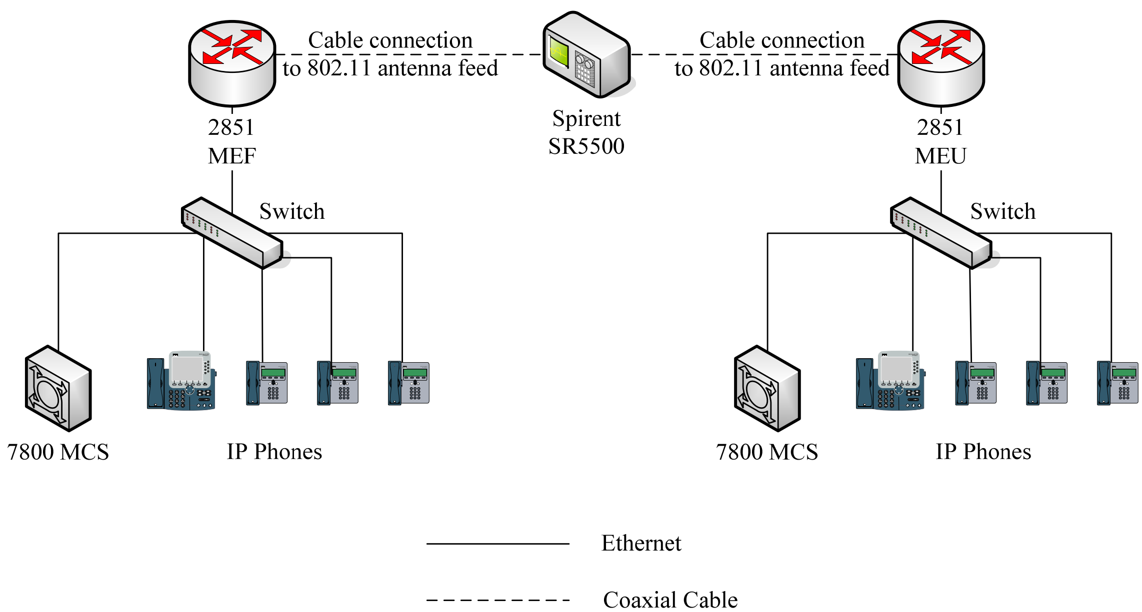

Figure 50. Suggested Testbed Alterations for Spirent SR5500 Connection to Cisco 2851 Router IEEE 802.11 Interface

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A

Useful IP Phone Information

All phones within the testbed have a web interface. A user can navigate to this page by typing the target IP phone's address into a browser. Figure 51 shows the initial page that opens for the target device.

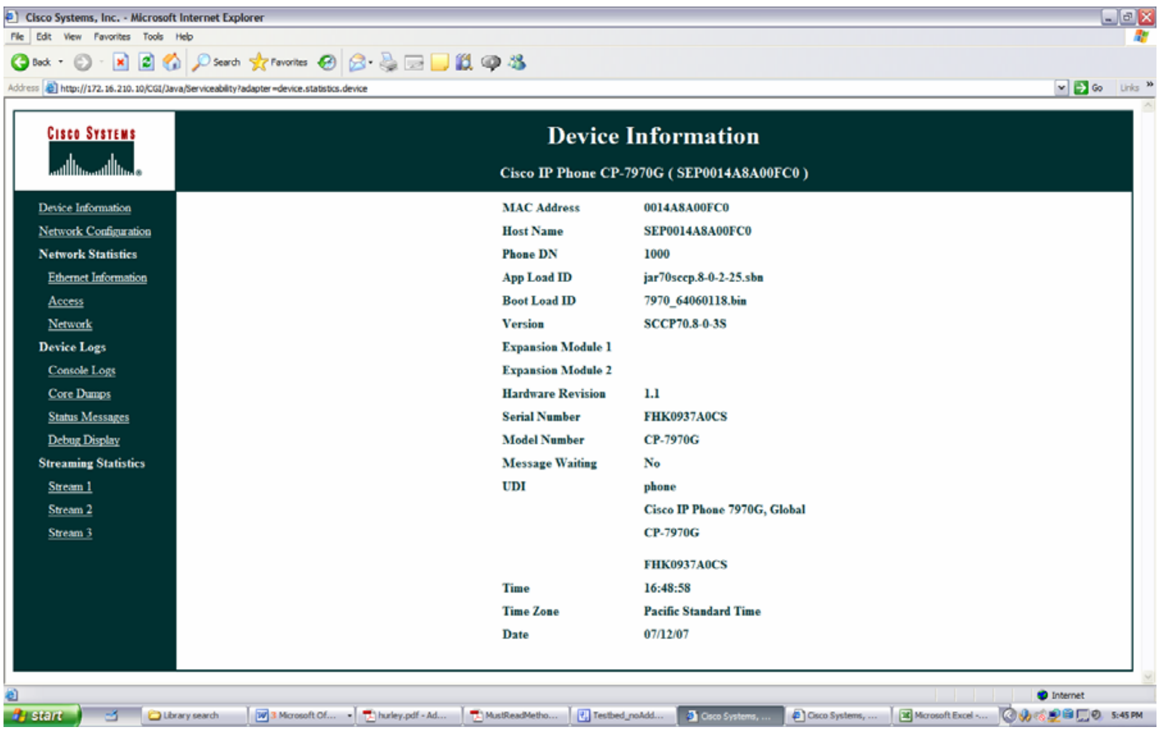

Figure 51. IP Phone Web Page

A wide variety of data from the previous three voice streams connected to this device are maintained under the Streaming Statistics group of the phone homepage. Figure 52 breaks out available items and their description as defined in [46]. The most current stream data is available for direct view on 7970G screens by pressing the ? button twice during an active call. Web displayed statistics can be exported to a Microsoft Excel spreadsheet by selecting the export link provided on the page.

Item	Description
Domain	Domain of the phone
Remote Address	IP address of the destination of the stream
Local Address	IP address of the phone
Sender Joins	Number of times the phone has started transmitting a stream
Receiver Joins	Number of times the phone has started receiving a stream
Byes	Number of times the phone has stopped transmitting a stream
Start Time	Internal time stamp indicating when Cisco CallManager requested that the phone start transmitting packets
Row Status	Whether the phone is streaming
Host Name	Host name of the phone
Sender Packets	Total number of packets sent by the phone
Sender Octets	Total number of octets sent by the phone
Sender Tool	Type of audio encoding used for the stream
Sender Reports	Number of times this streaming statistics report has been accessed from the web page (resets when the phone resets)
Sender Report Time	Internal time stamp indicating when this streaming statistics report was generated
Sender Start Time	Time that the stream started
Rcvr Lost Packets	Total number of packets lost
Rcvr Jitter	Maximum jitter of stream
Receiver Tool	Type of audio encoding used for the stream
Rcvr Reports	Number of times this streaming statistics report has been accessed from the web page (resets when the phone resets)
Rcvr Report Time	Internal time stamp indicating when this streaming statistics report was generated
Rcvr Packets	Total number of packets received by the phone
Rcvr Octets	Total number of octets received by the phone
Rcvr Start Time	Internal time stamp indicating when Cisco CallManager requested that the phone start receiving packets

Figure 52. Streaming Statistics Description (after [46])

The phones terminals can be unlocked to alter settings by pressing **#.

APPENDIX B

Cisco CallManager 5.0(4) Settings and Tips

All alterations to the testbed CallManager settings are in accordance with [38]. This appendix provides a general overview of some typical tasks used during testbed experiments and management. Further documentation and current recommended practices are available from the Cisco Systems web page [www.cisco.com]. The remainder of this appendix is organized into the following task sections:

- Login to testbed CallManagers
- Codec selection
- Music on Hold interface
- Adding/removing phone services
- Directory numbers
- Gateway management
- Dial patterns

Login to testbed CallManagers:

In order to access a CallManager web interface, a computer must have a valid IP address associated with the physical attachment to the testbed (i.e., 170.16.210.5 while attached to the switch on the MEF side of the network). Login is accomplished through the following steps:

- Open a web browser and search for the target CallManager IP address.
- Type CCMAdministrator and the current password when prompted.

Figure 53 shows the first page users encounter following a successful login sequence.

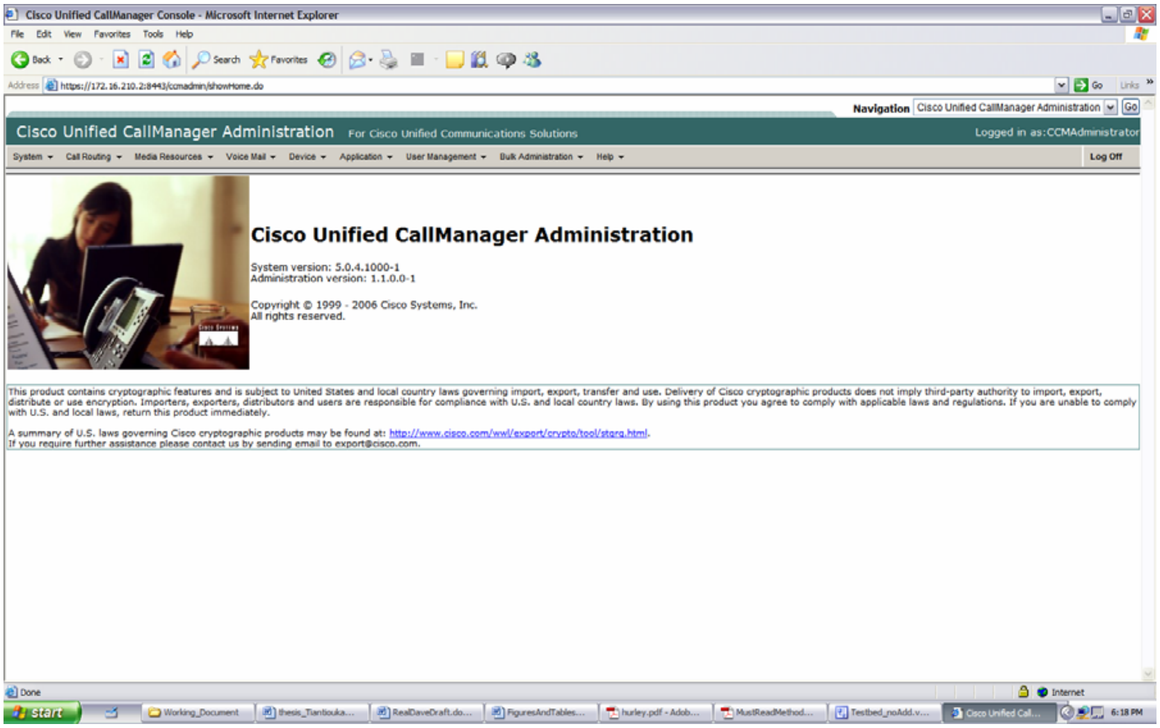

Figure 53. CallManager Login

Codec selection:

Screen shots of the following steps to select a codec are shown in Figure 54:

- From the "Systems" menu, select "Region",
- Select the region titled "Default,"
- Select the "Default" region in the window titled, "Modify Relationship to other Regions" (bottom left side of screen),
- Select the desired codec from the pull down menu titled, "Audio Codec" (bottom center of screen),
- Select the "Save" or "Cancel" button as appropriate, and
- If prompted, select the "Reset" button to implement changes across the testbed.

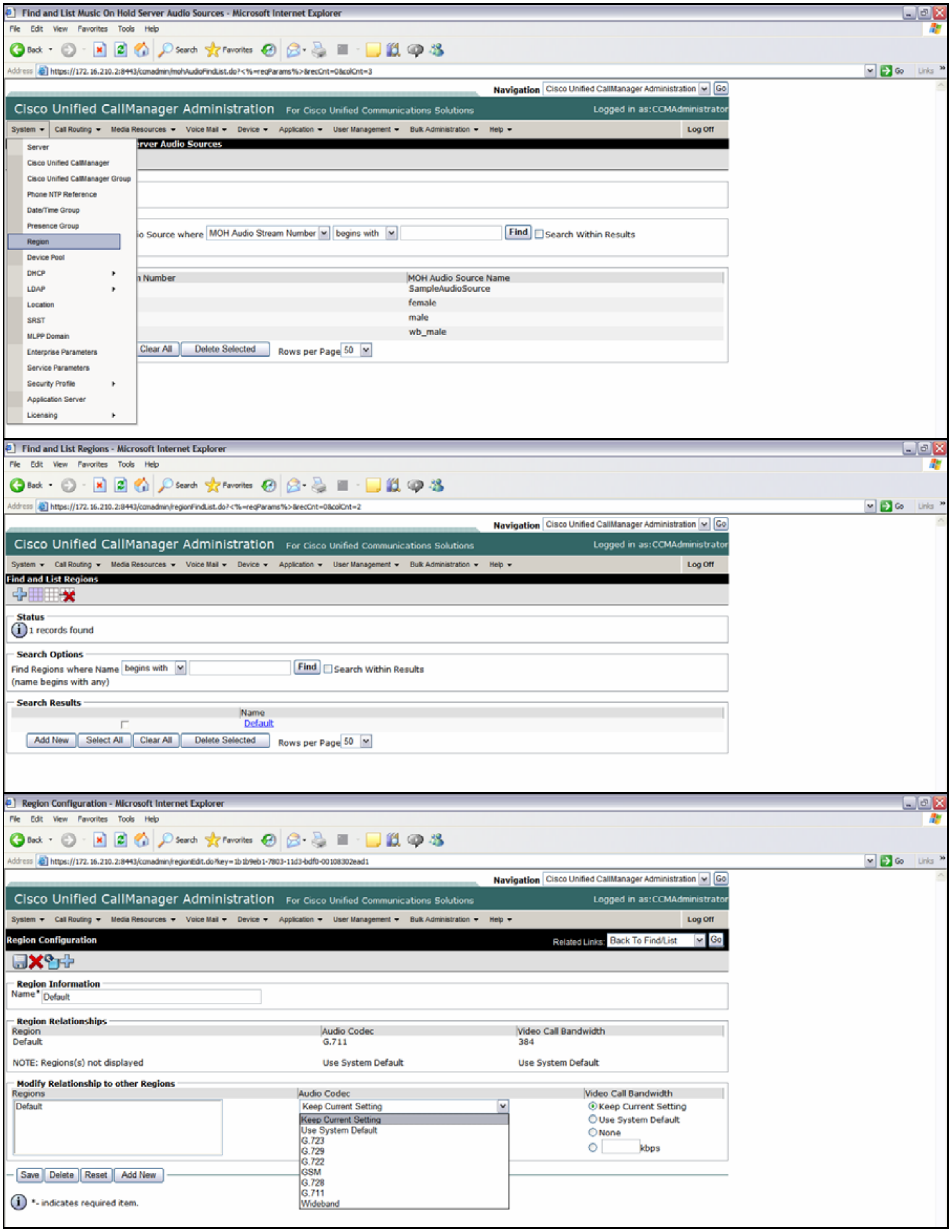

Figure 54. CallManager Codec Selection

Music On Hold (MOH) interface:

The Cisco 7800 series MCS on the MEF side of the testbed is configured to provide MOH server services. An MOH server stores the WAV files used for testbed experiments. Figures 55 – 58 provide screen shots of the steps required to add a WAV file to the testbed:

- From the “Systems” menu, select “Service Parameters,”
- In the “Server*” window, select the active MOH server IP address,
- In the “Service*” window, select “Cisco IP Voice Streaming Media App (Active)” from the pull down list,
- Scroll down and select the “Advanced” button,
- Highlight all codecs of interest in the “Supported MOH Codecs” section,
- Set the “Default MOH Volume Level” to 0,
- Select the “Save” button,
- From the “Media Resources” menu, select “Music On Hold Audio Source,”
- Select the “Add New” button to browse for file to upload, and
- Associate a free audio source number with the new file.

Users can assign MOH files to a designated phone by following the adding/removing phone services steps, outlined in the next section of this appendix.

Figure 55. CallManager Service Parameters Control

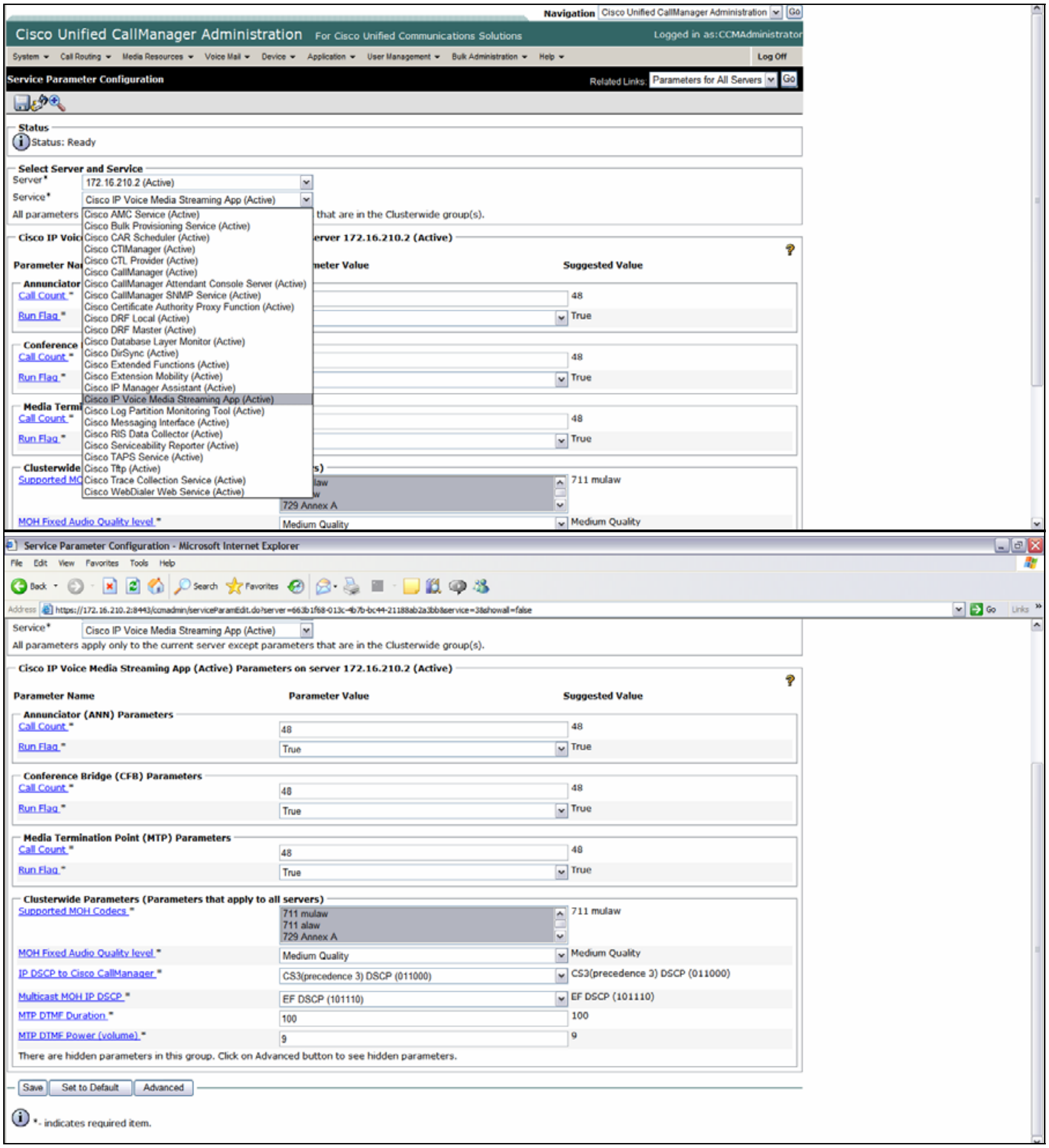

Figure 56. CallManager Streaming Media Application

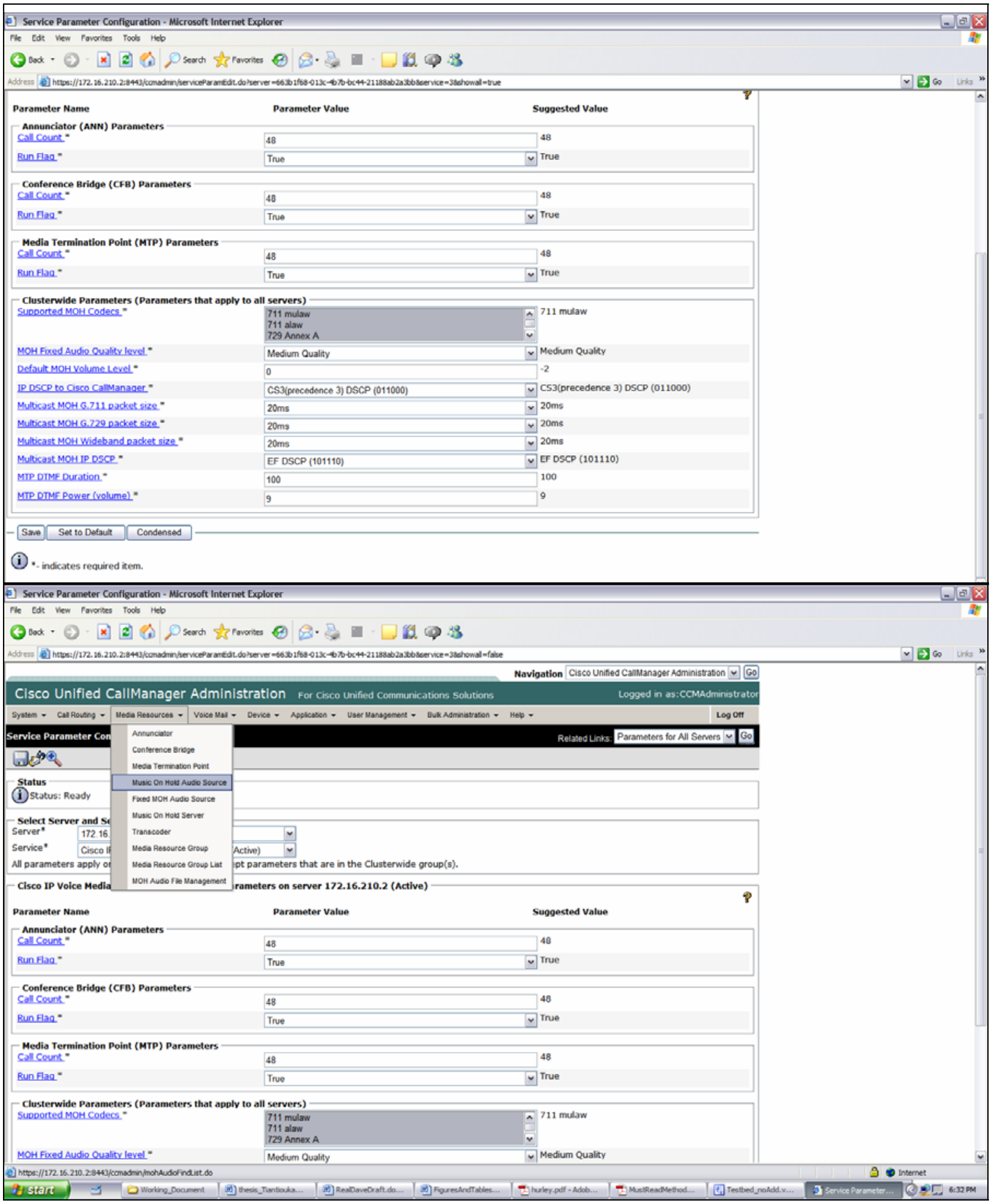

Figure 57. MOH Audio Source Settings

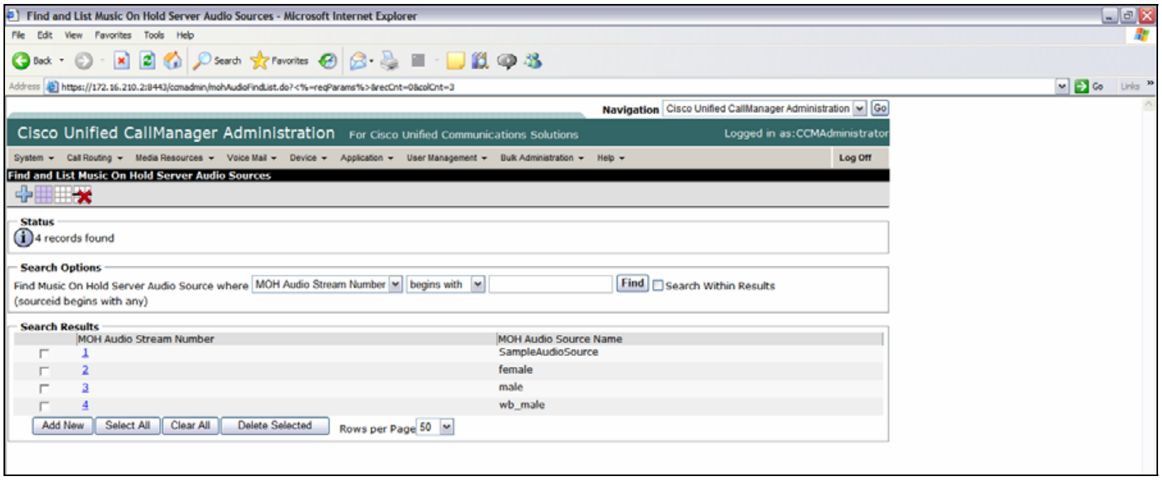

Figure 58. MOH Audio Stream Number Assignment

Adding/removing phone services:

The testbed auto discovery and assignment of device IP addresses has been disabled. This allows users to assign directory numbers to terminal devices according to dial plans of the experiment. The command sequence listed below describes the steps necessary to add/remove testbed IP phones, or to configure a specific MOH audio file to play when the selected terminal initiates a hold session. Figures 59 shows screen shots of these commands.

- From the “Device” menu, select “Phone,” then
- Select the “Find” button.

To add/delete phones:

- Select the “Add New” or “Delete Selected” button accordingly.

(or)

To modify an existing phone’s MOH source and directory number:

- Select the desired registered phone to edit,
- Assign a “User Hold MOH Audio Source” from the pull down menu in the “Device Information” window, and
- Assign an available directory to the phone number using hyperlinks in the “Association Information” window.

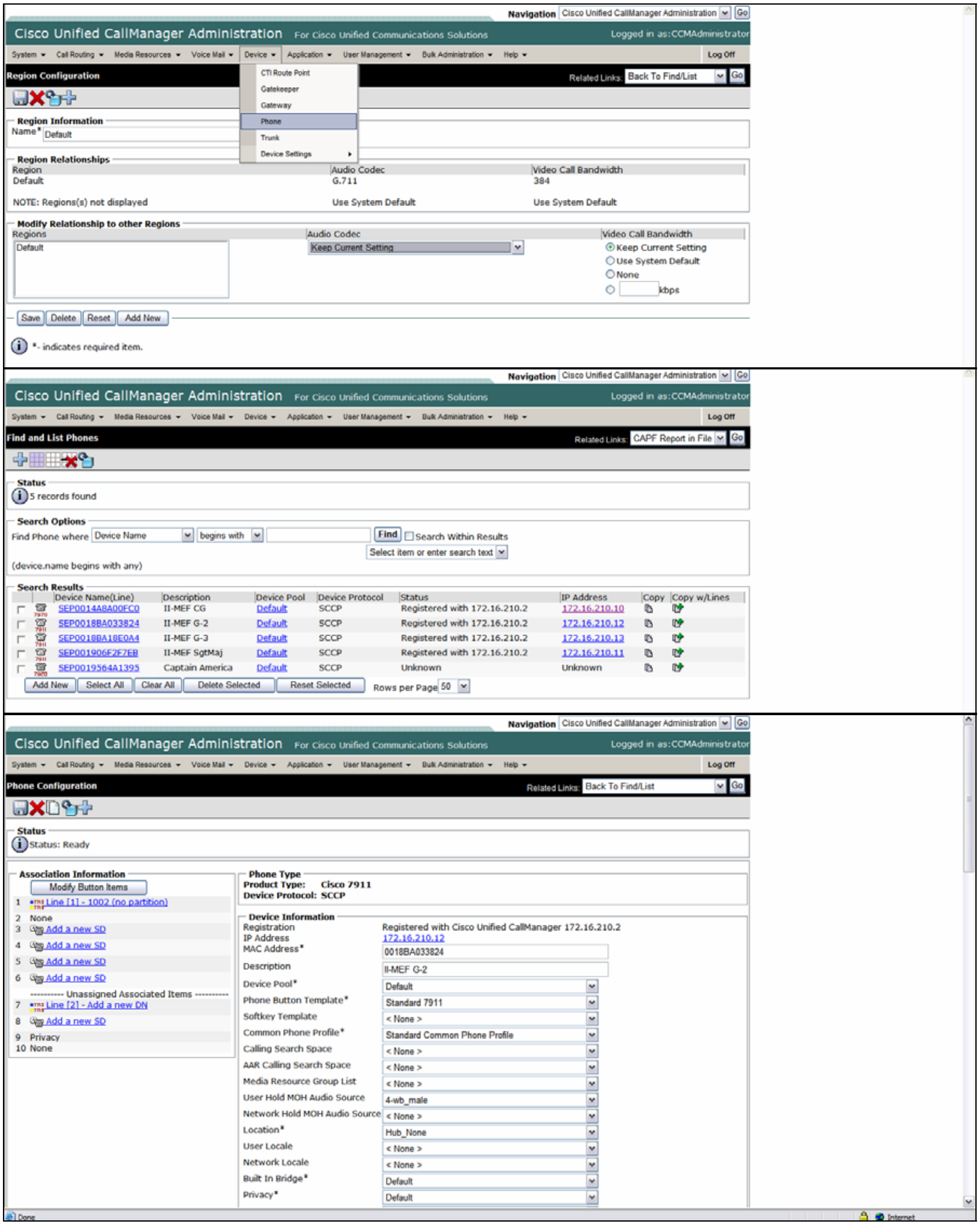

Figure 59. CallManager Phone Device Windows

Directory numbers:

Users can review the current list of directory numbers by browsing to the CallManager configuration page illustrated in Figure 60:

- From the "Call Routing" menu, select "Directory Number."

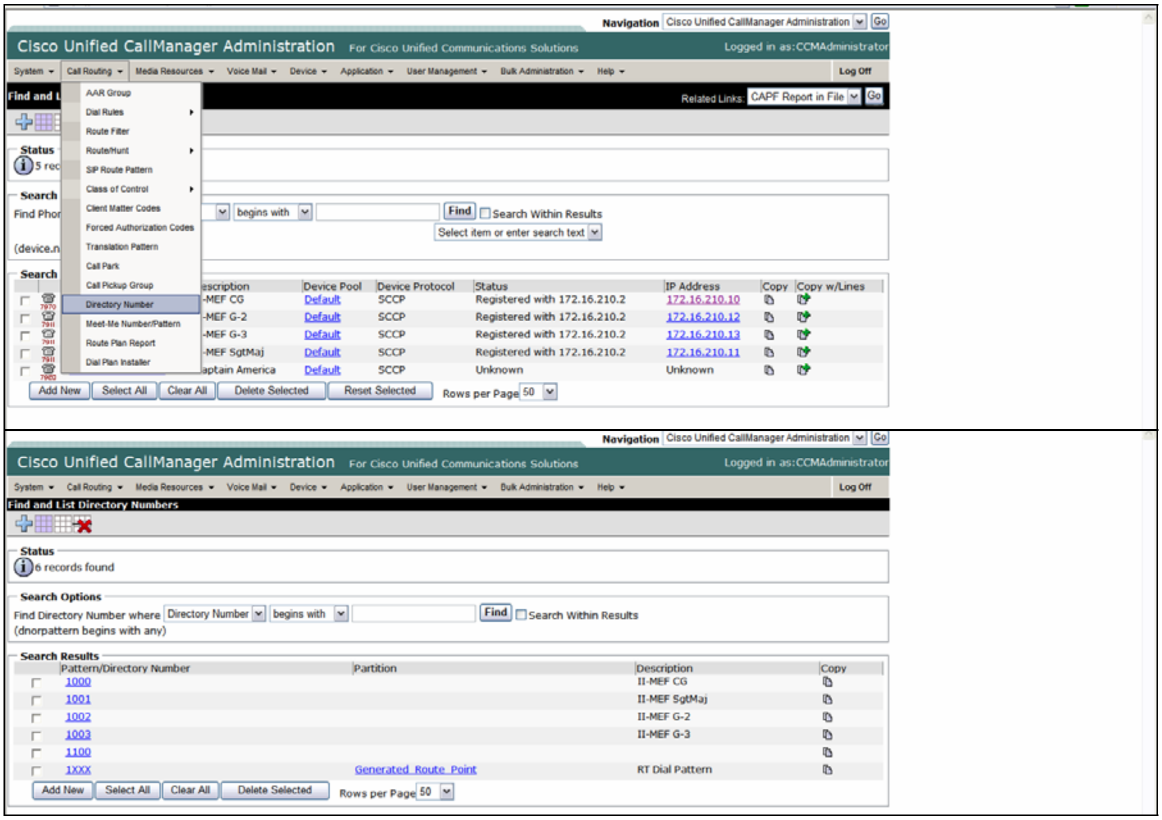

Figure 60. CallManager Directory

Gateway management:

Gateways are configured at two levels. Router command line interface inputs build the appropriate configuration file. Reference [47] provides instructions on gateway configuration. After the configuration file is loaded to the gateway, it must be registered within the CallManager software. This section will show the CallManager related items only. Figure 61 depicts the steps required to associate a gateway with the CallManager software. The testbed has one associated gateway identified by the current IP address assigned to the MEF 2851 interface connected to the MEFfiber 7200 router. In the event of network address adjustment or topology alterations, the gateway device name must be corrected using the following commands:

- From the "Device" menu, select "Gateway,"
- Type the IP address into the "Device Name*" field,
- Select the "Save" button, and
- If prompted, select the "Reset" button.

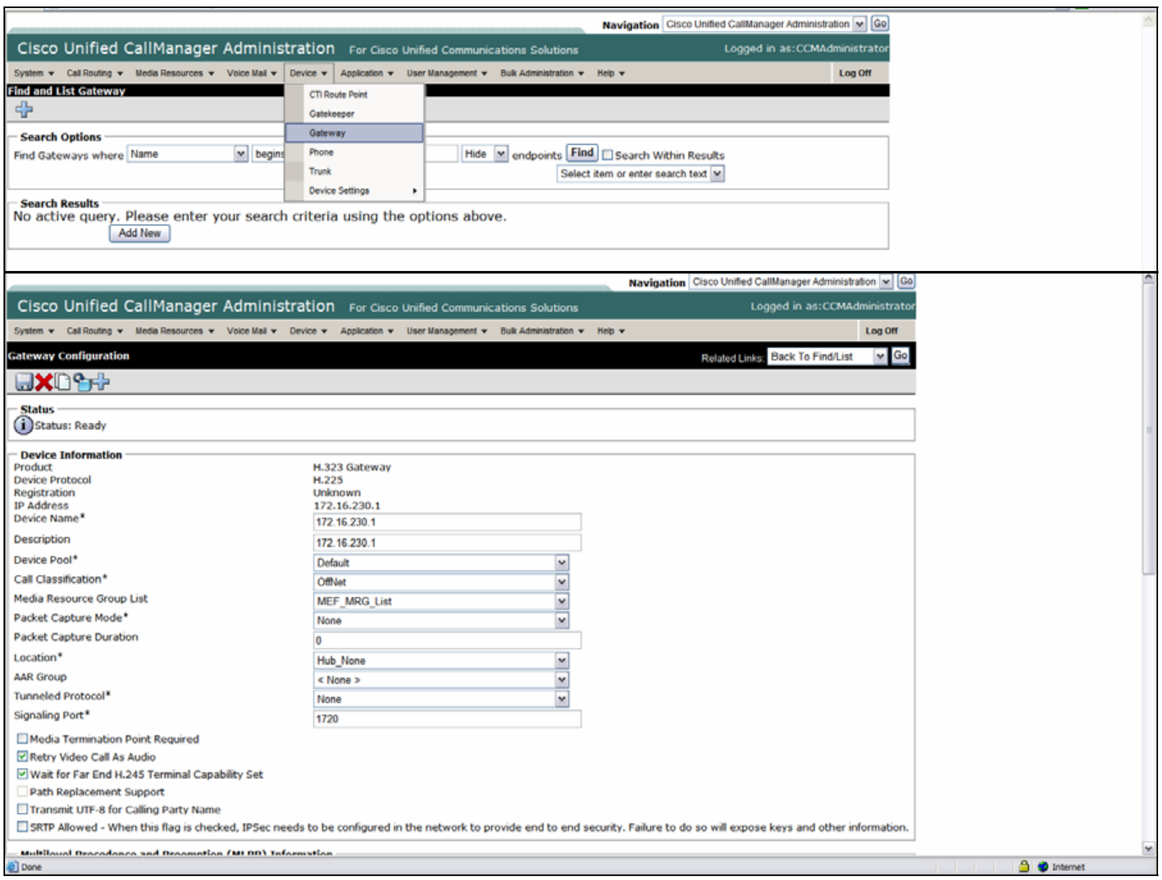

Figure 61. CallManager Gateway Configuration

Route patterns:

Route patterns link a sequence of dialed numbers to a specific call processing action. Current patterns are associated to the registered gateway for on/off net identification. On net number patterns receive an internal dial tone. Internal cluster calls are managed locally though a single CallManager. Off net number patterns receive an outside dial tone. Calls to/from terminals external to the cluster require signaling between CallManager units. In both cases, the route pattern is associated to the IP address of the gateway as shown in Figure 62. The following commands are provided to associate a route pattern to the existing gateway:

- From the "Call Routing" menu, select "Route/Hunt" and the submenu option "Route Pattern."
- Select a desired pattern to associate to the gateway, and
- Ensure the pattern registers the gateway IP address under the "Associated Devices" column when complete.

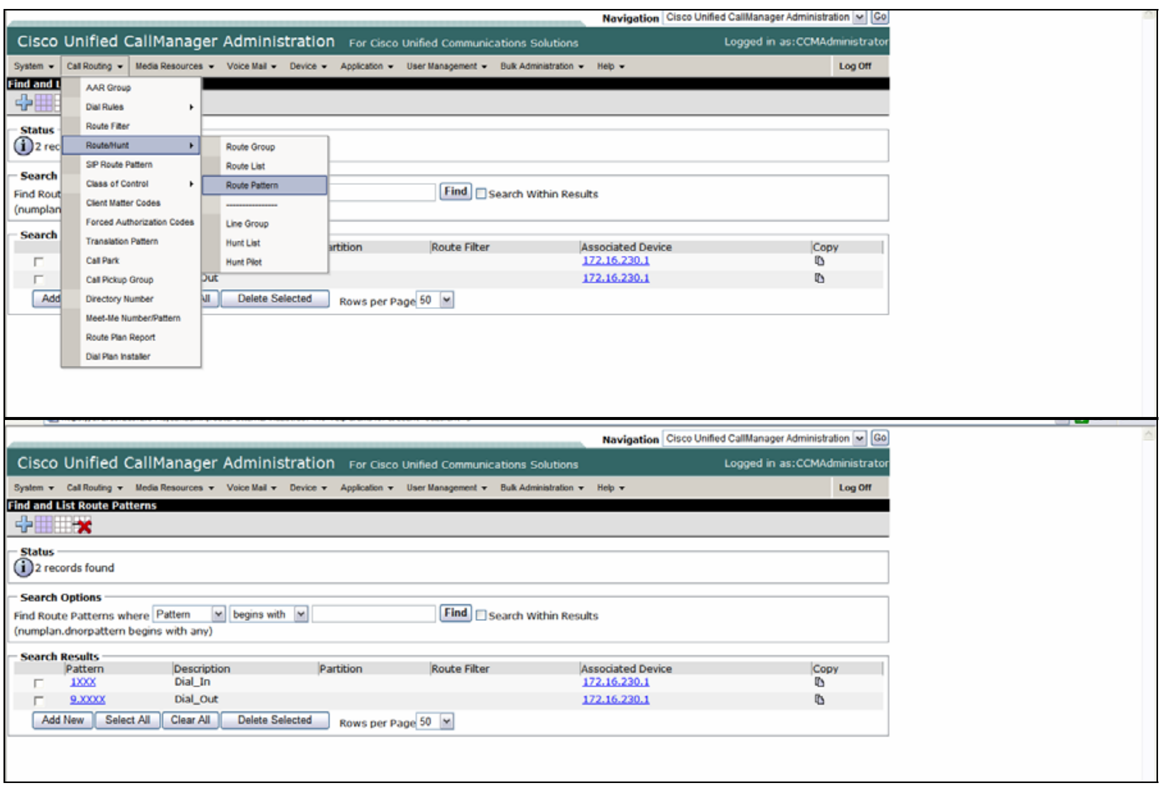

Figure 62. CallManager Route Pattern Configuration

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] T. S. Rappaport, *Wireless Communications Principles and Practice*, Second Edition, Pearson Prentice Hall, Upper Saddle River, 2002.

[2] J. E. Nuechterlein, P. J. Weiser, *Digital Crossroads*, The MIT Press, Cambridge, 2005.

[3] Marine Corps Combat Development Command, *Marine Corps Operating Concepts for a Changing Security Environment*, Second Edition, United States Marine Corps, Quantico, June 2007.

[4] J. Zhang, D. Yang, and Z. Quan, "Voice Quality of VoIP in Mobile Communications Systems," *2006 IEEE Radio and Wireless Symposium*, pp: 131 – 134, January 2006.

[5] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," February 2001.

[6] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality," August 1996.

[7] E. E. Zurek, J. Leffew, and W. A. Moreno, "Objective Evaluation of Voice Clarity Measurements for VoIP Compression Algorithms," *Proceedings of the Fourth IEEE International Caracas Conference on Devices, Circuits and Systems*, pp: T033-1 – T033-6, Aruba, April 2002.

[8] C.M. Chemick, et al., "Testing the Ability of Speech Recognizers to Measure the Effectiveness of Encoding Algorithms for Digital Speech Transmission," *IEEE Military Communications Conference, MILCOM 1999.*Volume 2, pp: 1468 – 1472, Atlantic City, November 1999.

[9] N. Tiantioukas, "Effects of the Wireless Channel, Signal Compression and Network Architecture on Speech Quality in VoIP Networks," Master's thesis, Naval Postgraduate School, Monterey California, 2007.

[10] F. J. Lawas-Grodek, et al., "SCPS-TP, TCP and Rate-Based Protocol Evaluation for High Delay, Error Prone Links," NASA Glenn Research Center, October 2002, Retrieved July 2007 from http://roland.grc.nasa.gov/~ivancic/papers_presentations/2002/T1020.pdf.

[11] B. Goode, "Voice Over Internet Protocol," *Proceedings of the IEEE,* Vol. 90, Issue 9, pp. 1495-1517, September 2002.

[12] J. Rosenburg, et. al., “Session Initiation Protocol,” IETF RFC 3261, June 2002.

[13] Y. Zhang, “SIP based VoIP network and its internetworking with the PSTN,” *Electronics and Communications Engineering Journal*, Vol. 14, Issue 6, pp. 273-282, December 2002.

[14] ITU-T Recommendation H.323, “Packet-based Multimedia Communications Systems,” June 2006.

[15] The Internet Engineering Consortium, “H. 323”, Web ProForum Tutorials, Retrieved July 2007 from http://www.iec.org.

[16] ITU-T Recommendation H.225.0, “Call Signalling Protocols and Media Stream Packetization for Packet-based Multimedia Communication,” May 2006.

[17] ITU-T Recommendation Q.931, “ISDN user-network interface layer 3 specification for basic call control,” May 1998.

[18] Cisco Systems, “Cisco Unified Communications Solution Reference Network Design (SRND) Based on Cisco CallManager 5.X,” February 2007, Retrieved July 2007 from www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a00806492bb.html.

[19] Cisco Systems, “Understanding H.323 Gatekeepers,” July 2006, Retrieved July 2007 from www.cisco.com/warp/public/788/voip/understand-gatekeepers.html.

[20] ITU-T Recommendation H.245, “Control Protocol for Multimedia Communication,” August 2005.

[21] R. L. Freeman, *Telecommunication Systems Engineering*, Wiley and Sons, Hoboken, 2004.

[22] ITU-T Recommendation G.711, “Pulse Code Modulation (PCM) of Voice Frequencies,” November 1988.

[23] ITU-T Recommendation G.729, “Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (AS-CELP),” January 2007.

[24] D. Chairporn, “Performance Evaluation of Voice over Internet Protocol,” Master’s thesis, Naval Postgraduate School, Monterey California, 2002.

[25] H. Schilzrinne, et al., “RTP: A Transport Protocol for Real-Time Applications,” IETF RFC 3550, July 2003.

[26] W. Stallings, *Data and Computer Communications*, Pearson Prentice Hall, Upper Saddle River, 2007.

[27] ITU-T Recommendation H.450.1, “Generic Functional Protocol for the Support of Supplementary Services in H.323,” February 1998.

[28] The Internet Engineering Consortium, “Voice Quality (VQ) in Converging Telephony and IP Networks”, Web ProForum Tutorials, Retrieved July 2007 from http://www.iec.org.

[29] ITU-T Recommendation G.114, “One-way Transmission Time,” May 2003.

[30] Cisco Systems, “Understand Delay in Packet Voice Networks,” March 2005, Retrieved July 2007 from http://www.cisco.com/warp/public/788/voip/delay-details.html.

[31] ITU-T Recommendation G.131, “’Talker Echo and Its Control,” November 2003.

[32] ITU-T Recommendation P.563, “’Single-ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications,” May 2004.

[33] ITU-T Recommendation G.107, “’The E-model, a Computational Model for Use in Transmission Planning,” March 2005.

[34] R. J. Deller, et al., *Discrete Time Processing of Speech Signals*, McMillian Press, London, 1993.

[35] W. Farrar, et al., “Wireless Communications Infrastructure in a Military Environment,” *Advances in Wired and Wireless Communication 2004, IEEE Sarnoff Symposium*, pp:103 – 106, Apr 2004.

[36] Cisco Systems, “Unified IP Phone 7970G Introduction,” Retrieved July 2007 from www.cisco.com/en/US/products/hw/phones/ps379/ps5440/index.html.

[37] Cisco Systems, “Unified IP Phone 7911G Introduction,” Retrieved July 2007 from www.cisco.com/en/US/products/ps6565/index.html.

[38] Cisco Systems, “Cisco Unified CallManager Administration Guide, Release 5.0(4),” Maintain and Operate Guides, 2006, Retrieved July 2007 from www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html.

[39] Speex Development Team, “Speex Samples”, Retrieved July 2007 from www.speex.org/samples.

[40] Netgear, “FS752TPS Software Admin Guide,” April 2006, Retrieved July 2007 from http://kbserver.netgear.com/products/FS752TPS.asp.

[41] ITU-T Recommendations H.221, "Frame structure for a 64 to 1920 kbit/s channel in audiovisual teleservices," March 2004.

[42] Adtech Incorporated, "Operating Maunal SX/14 Data Channel Simulator Version 2.0e", Honolulu, October 1995.

[43] Wireshark, "Network protocol analyzer," Retrieved July 2007 from www.wireshark.org.

[44] M. Montoro, "Cain & Abel v4.9.1," Retrieved July 2007 from www.oxid.it/cain.html.

[45] Nuance Communications, "Nuance Recognizer White Paper," January 2007, Retrieved July 2007 from www.nuance.com/recognizer/whitepapers.

[46] Cisco Systems, "Cisco Unified IP 7970 Series Administration Guide for Cisco Unified CallManager, Release 5.0 (for models 7970G and 7971G-GE)(SCCP)," Maintain and Operate Guides, February 2007, Retrieved July 2007 from www.cisco.com/en/US/products/hw/phones/ps379/products_administration_guide_book09186a00805ea051.html.

[47] Cisco Systems, "Cisco IOS H.323 Configuration Guide," June 2006, Retrieved July 2007 from www.cisco.com/en/US/products/sw/iosswrel/ps5207/products_configuration_guide_book09186a00801fcee1.html.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Marine Corps Representative
   Naval Postgraduate School
   Monterey, California

4. Director, Training and Education, MCCDC, Code C46
   Quantico, Virginia

5. Director Marine Corps Research Center, MCCDC, Code C40RC
   Quantico, Virginia

6. Marine Corps Tactical Systems Support Activity (Attn: Operations Officer)
   Camp Pendleton, California